Certificate of Mailing

Date of Deposit October 10 2001 Label Number: EL874187534US

I hereby certify under 37 C.F.R. § 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office to Addressee"** with sufficient postage on the date indicated above and is addressed to BOX PATENT APPLICATION, Assistant Commissioner for Patents, Washington, D.C. 20231.

Guy Beardsley
Printed name of person mailing correspondence

Guy Beardsley
Signature of person mailing correspondence

APPLICATION

FOR

UNITED STATES LETTERS PATENT

APPLICANT : Frederick M. Ausubel and Laurence G. Rahme
TITLE : VIRULENCE-ASSOCIATED NUCLEIC ACID SEQUENCES AND USES THEREOF

VIRULENCE-ASSOCIATED NUCLEIC ACID SEQUENCES AND USES THEREOF

Cross Reference to Related Applications

This application is a divisional of U.S. utility application 09/199,637, filed November 25, 1998, which is still pending and which claims benefit of U.S. provisional application serial number 60/066,517, filed November 25, 1997.

Background of the Invention

This invention relates to nucleic acid molecules, genes, and polypeptides that are related to microbial pathogenicity.

Pathogens employ a number of genetic strategies to cause infection and, occasionally, disease in their hosts. The expression of microbial pathogenicity is dependent upon complex genetic regulatory circuits. Knowledge of the themes in microbial pathogenicity is necessary for understanding pathogen virulence mechanisms and for the development of new "anti-virulence or anti-pathogenic" agents, which are needed to combat infection and disease.

In one particular example, the opportunistic human pathogen, *Pseudomonas aeruginosa,* is a ubiquitous gram-negative bacterium isolated from soil, water, and plants (Palleroni, J.N. In: *Bergey's Manual of Systematic Bacteriology*, ed., J.G. Holt, Williams & Wilkins, Baltimore, MD, pp. 141-172, 1984). A variety of *P. aeruginosa* virulence factors have been described and the majority of these, such as exotoxin A, elastase, and phospholipase C, were first detected biochemically on the basis of their cytotoxic activity (Fink, R.B., *Pseudomonas aeruginosa the Opportunist: Pathogenesis and Disease,*Boca Raton, CRC Press Inc., 1993). Subsequently, the genes corresponding to these factors or genes that regulate the expression of these factors were identified. In general, most pathogenicity-related genes in mammalian bacterial pathogens were first detected using a

bio-assay. In contrast to mammalian pathogens, simple systematic genetic strategies have been routinely employed to identify pathogenicity-related genes in plant pathogens. Following random transposon-mediated mutagenesis, thousands of mutant clones of the phytopathogen are inoculated separately into individual plants to determine if they contain a mutation that affects the pathogenic interaction with the host (Boucher et al., *J. Bacteriol.* 168:5626-5623, 1987; Comai and Kosuge, *J. Bacteriol.* 149:40-46, 1982; Lindgren et al., *J. Bacteriol.* 168:512-522, 1986; Rahme et al., *J. Bacteriol.* 173:575-586, 1991; Willis et al., *Mol. Plant-Microbe Interact.* 3:149-156, 1990). Comparable experiments using whole-animal mammalian pathogenicity models are not feasible because of the vast numbers of animals that must be subjected to pathogenic attack.

Summary of the Invention

We have identified and characterized a number of nucleic acid molecules, polypeptides, and small molecules (e.g., phenazines) that are involved in conferring pathogenicity and virulence to a pathogen. This discovery therefore provides a basis for drug-screening assays aimed at evaluating and identifying "anti-virulence" agents which are capable of blocking pathogenicity and virulence of a pathogen, e.g., by selectively switching pathogen gene expression on or off, or which inactivate or inhibit the activity of a polypeptide which is involved in the pathogenicity of a microbe. Drugs that target these molecules are useful as such anti-virulence agents.

In one aspect, the invention features an isolated nucleic acid molecule including a sequence substantially identical to any one of *BI48* (SEQ ID NO:1), ORF2 (SEQ ID NO:2), ORF3 (SEQ ID NO:4), ORF602c (SEQ ID NO:6), ORF214 (SEQ ID NO:8), ORF1242c (SEQ ID NO:10), ORF594 (SEQ ID NO:12), ORF1040 (SEQ ID NO:14), ORF1640c (SEQ ID NO:16), ORF2228c (SEQ ID NO:18), ORF2068c (SEQ ID NO:20), ORF1997 (SEQ ID NO:22), ORF2558c (SEQ ID NO:24), ORF2929c (SEQ ID NO:26), ORF3965c (SEQ ID NO:28), ORF3218 (SEQ ID NO:30), ORF3568 (SEQ ID

NO:32), ORF4506c (SEQ ID NO:34), ORF3973 (SEQ ID NO:36), ORF4271 (SEQ ID NO:38), ORF4698 (SEQ ID NO:40), ORF5028 (SEQ ID NO:42), ORF5080 (SEQ ID NO:44), ORF6479c (SEQ ID NO:46), ORF5496 (SEQ ID NO:48), ORF5840 (SEQ ID NO:50), ORF5899 (SEQ ID NO:52), ORF6325 (SEQ ID NO:54), ORF7567c (SEQ ID NO:56), ORF7180 (SEQ ID NO:58), ORF7501 (SEQ ID NO:60), ORF7584 (SEQ ID NO:62), ORF8208c (SEQ ID NO:64), ORF8109 (SEQ ID NO:66), ORF9005c (SEQ ID NO:68), ORF8222 (SEQ ID NO:70), ORF8755c (SEQ ID NO:72), ORF9431c (SEQ ID NO:74), ORF9158 (SEQ ID NO:76), ORF10125c (SEQ ID NO:78), ORF9770 (SEQ ID NO:80), ORF9991 (SEQ ID NO:82), ORF10765c (SEQ ID NO:84), ORF10475 (SEQ ID NO:86), ORF11095c (SEQ ID NO:88), ORF11264 (SEQ ID NO:90), ORF11738 (SEQ ID NO:92), ORF12348c (SEQ ID NO:94), ORF12314c (SEQ ID NO:96), ORF13156c (SEQ ID NO:98), ORF12795 (SEQ ID NO:100), ORF13755c (SEQ ID NO:210), ORF13795c (SEQ ID NO:212), ORF14727c (SEQ ID NO:214), ORF13779 (SEQ ID NO:216), ORF14293c (SEQ ID NO:218), ORF14155 (SEQ ID NO:220), ORF14360 (SEQ ID NO:222), ORF15342c (SEQ ID NO:224), ORF15260c (SEQ ID NO:226), ORF14991 (SEQ ID NO:228), ORF15590c (SEQ ID NO:230), ORF15675c (SEQ ID NO:232), ORF16405 (SEQ ID NO:234), ORF16925 (SEQ ID NO:236), ORF17793c (SEQ ID NO:238), ORF18548c (SEQ ID NO:240), ORF17875 (SEQ ID NO:242), ORF18479 (SEQ ID NO:244), ORF19027c (SEQ ID NO:246), ORF19305 (SEQ ID NO:248), ORF19519 (SEQ ID NO:250), ORF19544 (SEQ ID NO:252), ORF20008 (SEQ ID NO:254), ORF20623c (SEQ ID NO:256), ORF21210c (SEQ ID NO:258), ORF21493c (SEQ ID NO:260), ORF21333 (SEQ ID NO:262), ORF22074c (SEQ ID NO:264), ORF21421 (SEQ ID NO:266), ORF22608c (SEQ ID NO:268), ORF22626 (SEQ ID NO:270), ORF23228 (SEQ ID NO:272),ORF23367 (SEQ ID NO:274), ORF25103c (SEQ ID NO:276), ORF23556 (SEQ ID NO:278), ORF26191c (SEQ ID NO:280), ORF23751 (SEQ ID NO:282), ORF24222 (SEQ ID NO:284), ORF24368 (SEQ ID NO:286), ORF24888c (SEQ ID NO:288), ORF25398c (SEQ ID NO:290),

ORF25892c (SEQ ID NO:292), ORF25110 (SEQ ID NO:294), ORF25510 (SEQ ID NO:296), ORF26762c (SEQ ID NO:298), ORF26257 (SEQ ID NO:300), ORF26844c (SEQ ID NO:302), ORF26486 (SEQ ID NO:304), ORF26857c (SEQ ID NO:306), ORF27314c (SEQ ID NO:308), ORF27730c (SEQ ID NO:310), ORF26983 (SEQ ID NO:312), ORF28068c (SEQ ID NO:314), ORF27522 (SEQ ID NO:316), ORF28033c (SEQ ID NO:318), ORF29701c (SEQ ID NO:320), ORF28118 (SEQ ID NO:322), ORF28129 (SEQ ID NO:324), ORF29709c (SEQ ID NO:326), ORF29189 (SEQ ID NO:328), ORF29382 (SEQ ID NO:330), ORF30590c (SEQ ID NO:332), ORF29729 (SEQ ID NO:334), ORF30221 (SEQ ID NO:336), ORF30736c (SEQ ID NO:338), ORF30539 (SEQ ID NO:340), ORF31247c (SEQ ID NO:342), ORF31539c (SEQ ID NO:346), ORF31222 (SEQ ID NO:348), ORF31266 (SEQ ID NO:350), ORF31661c (SEQ ID NO:352), ORF32061c (SEQ ID NO:354), ORF32072c (SEQ ID NO:356), ORF31784 (SEQ ID NO:358), ORF32568c (SEQ ID NO:360), ORF33157c (SEQ ID NO:362), ORF32539 (SEQ ID NO:364), ORF33705c (SEQ ID NO:366), ORF32832 (SEQ ID NO:368), ORF33547c (SEQ ID NO:370), ORF33205 (SEQ ID NO:372), ORF33512 (SEQ ID NO:374), ORF33771 (SEQ ID NO:376), ORF34385c (SEQ ID NO:378), ORF33988 (SEQ ID NO:380), ORF34274 (SEQ ID NO:382), ORF34726c (SEQ ID NO:384), ORF34916 (SEQ ID NO:386), ORF35464c (SEQ ID NO:388), ORF35289 (SEQ ID NO:390), ORF35410 (SEQ ID NO:392), ORF35907c (SEQ ID NO:394),ORF35534 (SEQ ID NO:396), ORF35930 (SEQ ID NO:398), ORF36246 (SEQ ID NO:400), ORF26640c (SEQ ID NO:402), ORF36739 (SEQ ID NO:404), ORF37932c (SEQ ID NO:406), ORF38640c (SEQ ID NO:408), ORF39309c (SEQ ID NO:410), ORF38768 (SEQ ID NO:412), ORF40047c (SEQ ID NO:414), ORF40560c (SEQ ID NO:416), ORF40238 (SEQ ID NO:418), ORF40329 (SEQ ID NO:420), ORF40709c (SEQ ID NO:422), ORF40507 (SEQ ID NO:424), ORF41275c (SEQ ID NO:426), ORF42234c (SEQ ID NO:428), ORF41764c (SEQ ID NO:430), ORF41284 (SEQ ID NO:432), ORF41598 (SEQ ID NO:434), ORF42172c (SEQ ID NO:436), ORF42233c

(SEQ ID NO:438), 33A9 (SEQ ID NO:102, 189, 190, 191, 192, 193, 194, 195, 196, 197, and 198), *34B12* (SEQ ID NO:104), *34B12-ORF1* (SEQ ID NO:105), *34B12-ORF2* (SEQ ID NO:106), *36A4* (SEQ ID NO:109), *36A4 contig* (SEQ ID NO:111), *23A2* (SEQ ID NO:112), *3E8 phn(-)*(SEQ ID NO:114), *3E8 contigPAO1* (SEQ ID NO:115), *34H4* (SEQ ID NO:118), *33C7* (SEQ ID NO:119), *25a12.3* (SEQ ID NO:120), *8C12* (SEQ ID NO:121), *2A8* (SEQ ID NO:122), *41A5* (SEQ ID NO:123), *50E12* (SEQ ID NO:124), *35A9* (SEQ ID NO:125), *pho23* (SEQ ID NO:126), *16G12* (SEQ ID NO:127), *25F1* (SEQ ID NO:128), *PA14 degP* (SEQ ID NO:131), 1126 contig (SEQ ID NO:135), contig 1344 (SEQ ID NO:136), ORFA (SEQ ID NO:440), ORFB (SEQ ID NO:441), ORFC (SEQ ID NO:442), *phzR* (SEQ ID NO:164, and 1G2 (SEQ ID NO:137). Preferably, the isolated nucleic acid molecule includes any of the above-described sequences or a fragment thereof; and is derived from a pathogen (e.g., from a bacterial pathogen such as *Pseudomonas aeruginosa*). Additionally, the invention includes a vector and a cell, each of which includes at least one of the isolated nucleic acid molecules of the invention; and a method of producing a recombinant polypeptide involving providing a cell transformed with a nucleic acid molecule of the invention positioned for expression in the cell, culturing the transformed cell under conditions for expressing the nucleic acid molecule, and isolating a recombinant polypeptide. The invention further features recombinant polypeptides produced by such expression of an isolated nucleic acid molecule of the invention, and substantially pure antibodies that specifically recognize and bind such a recombinant polypeptides.

In an another aspect, the invention features a substantially pure polypeptide including an amino acid sequence that is substantially identical to the amino acid sequence of any one of ORF2 (SEQ ID NO:3), ORF3 (SEQ ID NO:5), ORF602c (SEQ ID NO:7), ORF214 (SEQ ID NO:9),ORF1242c (SEQ ID NO:11), ORF594 (SEQ ID NO:13), ORF1040 (SEQ ID NO:15), ORF1640c (SEQ ID NO:17), ORF2228c (SEQ ID NO:19), ORF2068c (SEQ ID NO:21), ORF1997 (SEQ ID NO:23), ORF2558c (SEQ ID

NO:25), ORF2929c (SEQ ID NO:27), ORF3965c (SEQ ID NO:29), ORF3218 (SEQ ID NO:31), ORF3568 (SEQ ID NO:33), ORF4506c (SEQ ID NO:35), ORF3973 (SEQ ID NO:37), ORF4271 (SEQ ID NO:39), ORF4698 (SEQ ID NO:41), ORF5028 (SEQ ID NO:43), ORF5080 (SEQ ID NO:45), ORF6479c (SEQ ID NO:47), ORF5496 (SEQ ID NO:49), ORF5840 (SEQ ID NO:51), ORF5899 (SEQ ID NO:53), ORF6325 (SEQ ID NO:55), ORF7567c (SEQ ID NO:57), ORF7180 (SEQ ID NO:59), ORF7501 (SEQ ID NO:61), ORF7584 (SEQ ID NO:63), ORF8208c (SEQ ID NO:65), ORF8109 (SEQ ID NO:67), ORF9005c (SEQ ID NO:69), ORF8222 (SEQ ID NO:71), ORF8755c (SEQ ID NO:73), ORF9431c (SEQ ID NO:75), ORF9158 (SEQ ID NO:77), ORF10125c (SEQ ID NO:79), ORF9770 (SEQ ID NO:81), ORF9991 (SEQ ID NO:83), ORF10765c (SEQ ID NO:85), ORF10475 (SEQ ID NO:87), ORF11095c (SEQ ID NO:89), ORF11264 (SEQ ID NO:91), ORF11738 (SEQ ID NO:93), ORF12348c (SEQ ID NO:95), ORF12314c (SEQ ID NO:97), ORF13156c (SEQ ID NO:99), ORF12795 (SEQ ID NO:101), ORF13755c (SEQ ID NO:211), ORF13795c (SEQ ID NO:213), ORF14727c (SEQ ID NO:215), ORF13779 (SEQ ID NO:217), ORF14293c (SEQ ID NO:219), ORF14155 (SEQ ID NO:221), ORF14360 (SEQ ID NO:223), ORF15342c (SEQ ID NO:225), ORF15260c (SEQ ID NO:227), ORF14991 (SEQ ID NO:229), ORF15590c (SEQ ID NO:231), ORF15675c (SEQ ID NO:233), ORF16405 (SEQ ID NO:235), ORF16925 (SEQ ID NO:237), ORF17793c (SEQ ID NO:239), ORF18548c (SEQ ID NO:241), ORF17875 (SEQ ID NO:243), ORF18479 (SEQ ID NO:245), ORF19027c (SEQ ID NO:247), ORF19305 (SEQ ID NO:249), ORF19519 (SEQ ID NO:251), ORF19544 (SEQ ID NO:253), ORF20008 (SEQ ID NO:255), ORF20623c (SEQ ID NO:257), ORF21210c (SEQ ID NO:259), ORF21493c (SEQ ID NO:261), ORF21333 (SEQ ID NO:263), ORF22074c (SEQ ID NO:265), ORF21421 (SEQ ID NO:267), ORF22608c (SEQ ID NO:269), ORF22626 (SEQ ID NO:271), ORF23228 (SEQ ID NO:273), ORF23367 (SEQ ID NO:275), ORF25103c (SEQ ID NO:277), ORF23556 (SEQ ID NO:279), ORF26191c (SEQ ID NO:281), ORF23751 (SEQ ID NO:283), ORF24222 (SEQ ID NO:285),

ORF24368 (SEQ ID NO:287), ORF24888c (SEQ ID NO:289), ORF25398c (SEQ ID NO:291), ORF25892c (SEQ ID NO:293), ORF25110 (SEQ ID NO:295), ORF25510 (SEQ ID NO:297), ORF26762c (SEQ ID NO:299), ORF26257 (SEQ ID NO:301), ORF26844c (SEQ ID NO:303), ORF26486 (SEQ ID NO:305), ORF26857c (SEQ ID NO:307), ORF27314c (SEQ ID NO:309), ORF27730c (SEQ ID NO:311), ORF26983 (SEQ ID NO:313), ORF28068c (SEQ ID NO:315), ORF27522 (SEQ ID NO:317), ORF28033c (SEQ ID NO:319), ORF29701c (SEQ ID NO:321), ORF28118 (SEQ ID NO:323), ORF28129 (SEQ ID NO:325), ORF29709c (SEQ ID NO:327), ORF29189 (SEQ ID NO:329), ORF29382 (SEQ ID NO:331), ORF30590c (SEQ ID NO:333), ORF29729 (SEQ ID NO:335), ORF30221 (SEQ ID NO:337), ORF30736c (SEQ ID NO:339), ORF30539 (SEQ ID NO:341), ORF31247c (SEQ ID NO:343), ORF30963c (SEQ ID NO:345), ORF31539c (SEQ ID NO:347), ORF31222 (SEQ ID NO:349), ORF31266 (SEQ ID NO:351), ORF31661c (SEQ ID NO:353), ORF32061c (SEQ ID NO:355), ORF32072c (SEQ ID NO:357), ORF31784 (SEQ ID NO:359), ORF32568c (SEQ ID NO:361), ORF33157c (SEQ ID NO:363), ORF32530 (SEQ ID NO:365), ORF33705c (SEQ ID NO:367), ORF32832 (SEQ ID NO:369), ORF33547c (SEQ ID NO:371), ORF33205 (SEQ ID NO:373), ORF33512 (SEQ ID NO:375), ORF33771 (SEQ ID NO:377), ORF34385c (SEQ ID NO:379), ORF33988 (SEQ ID NO:381), ORF34274 (SEQ ID NO:383), ORF34726c (SEQ ID NO:385), ORF34916 (SEQ ID NO:387), ORF35464c (SEQ ID NO:389), ORF35289 (SEQ ID NO:391), ORF35410 (SEQ ID NO:393), ORF35907c (SEQ ID NO:395), ORF35534 (SEQ ID NO:397), ORF35930 (SEQ ID NO:399), ORF36246 (SEQ ID NO:401), ORF26640c (SEQ ID NO:403), ORF36769 (SEQ ID NO:405), ORF37932c (SEQ ID NO:407), ORF38640c (SEQ ID NO:409), ORF39309c (SEQ ID NO:411), ORF38768 (SEQ ID NO:413), ORF40047c (SEQ ID NO:415), ORF40560c (SEQ ID NO:417), ORF40238 (SEQ ID NO:419), ORF40329 (SEQ ID NO:421), ORF40709c (SEQ ID NO:423), ORF40507 (SEQ ID NO:425), ORF41275c (SEQ ID NO:427), ORF42234c (SEQ ID NO:429), ORF41764c

(SEQ ID NO:431), ORF41284 (SEQ ID NO:433), ORF41598 (SEQ ID NO:435), ORF42172c (SEQ ID NO:437), ORF42233c (SEQ ID NO:439), *33A9* (SEQ ID NOS:103, 199, 200, 201, 202, 203, 204, 205, 206, 207, and 208), *34B12-ORF1* (SEQ ID NO:107), *34B12-ORF2* (SEQ ID NO:108), *36A4* (SEQ ID NO:110), *3E8phzA* (SEQ ID NO:116), *3E8phzB* (SEQ ID NO:117), PhzR (SEQ ID NO:165), ORFA (SEQ ID NO:443), ORFB (SEQ ID NO:444), ORFC (SEQ ID NO:445), and *PA14 degP* (SEQ ID NO:132). Preferably, the substantially pure polypeptide includes any of the above-described sequences of a fragment thereof; and is derived from a pathogen (e.g., from a bacterial pathogen such as *Pseudomonas aeruginosa*).

In yet another related aspect, the invention features a method for identifying a compound which is capable of decreasing the expression of a pathogenic virulence factor (e.g., at the transcriptional or post-transcriptional levels), involving (a) providing a pathogenic cell expressing any one of the isolated nucleic acid molecules of the invention; and (b) contacting the pathogenic cell with a candidate compound, a decrease in expression of the nucleic acid molecule following contact with the candidate compound identifying a compound which decreases the expression of a pathogenic virulence factor. In preferred embodiments, the pathogenic cell infects a mammal (e.g., a human) or a plant.

In yet another related aspect, the invention features a method for identifying a compound which binds a polypeptide, involving (a) contacting a candidate compound with a substantially pure polypeptide including any one of the amino acid sequences of the invention under conditions that allow binding; and (b) detecting binding of the candidate compound to the polypeptide.

In addition, the invention features a method of treating a pathogenic infection in a mammal, involving (a) identifying a mammal having a pathogenic infection; and (b) administering to the mammal a therapeutically effective amount of a composition which inhibits the expression or activity of a polypeptide encoded by any one of the nucleic acid

molecules of the invention. In preferred embodiments, the pathogen is *Pseudomonas aeruginosa.*

In yet another aspect, the invention features a method of treating a pathogenic infection in a mammal, involving (a) identifying a mammal having a pathogenic infection; and (b) administering to the mammal a therapeutically effective amount of a composition which binds and inhibits a polypeptide encoded by any one of the amino acid sequences of the invention. In preferred embodiments, the pathogenic infection is caused by *Pseudomonas aeruginosa.*

Moreover, the invention features a method of identifying a compound which inhibits the virulence of a *Pseudomonas* cell, involving (a) providing a *Pseudomonas* cell; (b) contacting the cell with a candidate compound; and (c) detecting the presence of a phenazine, wherein a decrease in the phenazine relative to an untreated control culture is an indication that the compound inhibits the virulence of the *Pseudomonas* cell. In preferred embodiments, the cell is *Pseudomonas aeruginosa*; the cell is present in a cell culture; and the phenazine is detected by spectroscopy (e.g., pyocyanin is detected at an absorbance of 520 nm). Pyocyanin is generally detected according to any standard method, e.g., those described herein.

By "isolated nucleic acid molecule" is meant a nucleic acid (e.g., a DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term includes an RNA molecule which is transcribed from a DNA molecule, as well as a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By a "substantially pure polypeptide" is meant a polypeptide of the invention that has been separated from components which naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide of the invention. A substantially pure polypeptide of the invention may be obtained, for example, by extraction from a natural source (for example, a pathogen); by expression of a recombinant nucleic acid encoding such a polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 25% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). Preferably, such a sequence is at least 30%, 40%, 50%, 60%, more preferably 80%, and most preferably 90% or even 95% identical at the amino acid level or nucleic acid to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine,

arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

By "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) a polypeptide of the invention.

By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (i.e., facilitates the production of, for example, a recombinant polypeptide of the invention, or an RNA molecule).

By "purified antibody" is meant antibody which is at least 60%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 99%, by weight, antibody. A purified antibody of the invention may be obtained, for example, by affinity chromatography using a recombinantly-produced polypeptide of the invention and standard techniques.

By "specifically binds" is meant a compound or antibody which recognizes and binds a polypeptide of the invention but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample, which naturally includes a polypeptide of the invention.

By "derived from" is meant isolated from or having the sequence of a naturally-occurring sequence (e.g., a cDNA, genomic DNA, synthetic, or combination thereof).

By "inhibiting a pathogen" is meant the ability of a candidate compound to decrease, suppress, attenuate, diminish, or arrest the development or progression of a pathogen-mediated disease or an infection in a eukaryotic host organism. Preferably, such inhibition decreases pathogenicity by at least 5%, more preferably by at least 25%,

and most preferably by at least 50%, as compared to symptoms in the absence of the candidate compound in any appropriate pathogenicity assay (for example, those assays described herein). In one particular example, inhibition may be measured by monitoring pathogenic symptoms in a host organism exposed to a candidate compound or extract, a decrease in the level of symptoms relative to the level of pathogenic symptoms in a host organism not exposed to the compound indicating compound-mediated inhibition of the pathogen.

By "pathogenic virulence factor" is meant a cellular component (e.g., a protein such as a transcription factor, as well as the gene which encodes such a protein) without which the pathogen is incapable of causing disease or infection in a eukaryotic host organism.

The invention provides a number of targets that are useful for the development of drugs that specifically block the pathogenicity of a microbe. In addition, the methods of the invention provide a facile means to identify compounds that are safe for use in eukaryotic host organisms (i.e., compounds which do not adversely affect the normal development and physiology of the organism), and efficacious against pathogenic microbes (i.e., by suppressing the virulence of a pathogen). In addition, the methods of the invention provide a route for analyzing virtually any number of compounds for an anti-virulence effect with high-volume throughput, high sensitivity, and low complexity. The methods are also relatively inexpensive to perform and enable the analysis of small quantities of active substances found in either purified or crude extract form.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

Detailed Description

The drawings will first be described.

Drawings

Fig. 1 is a schematic diagram showing the physical map of cosmid BI48 (SEQ ID NO:1) and the orientation of the identified open reading frames (ORFs).

Fig. 2 shows the nucleotide sequence of cosmid BI48 (SEQ ID NO:1).

Fig. 3 shows the nucleotide sequences for ORF2 (SEQ ID NO:2), ORF3 (SEQ ID NO:4), ORF602c (SEQ ID NO:6), ORF214 (SEQ ID NO:8), ORF1242c (SEQ ID NO:10), ORF594 (SEQ ID NO:12), ORF1040 (SEQ ID NO:14), ORF1640c (SEQ ID NO:16), ORF2228c (SEQ ID NO:18), ORF2068c (SEQ ID NO:20), ORF1997 (SEQ ID NO:22), ORF2558c (SEQ ID NO:24), ORF2929c (SEQ ID NO:26), ORF3965c (SEQ ID NO:28), ORF3218 (SEQ ID NO:30), ORF3568 (SEQ ID NO:32), ORF4506c (SEQ ID NO:34), ORF3973 (SEQ ID NO:36), ORF4271 (SEQ ID NO:38), ORF4698 (SEQ ID NO:40), ORF5028 (SEQ ID NO:42), ORF5080 (SEQ ID NO:44), ORF6479c (SEQ ID NO:46), ORF5496 (SEQ ID NO:48), ORF5840 (SEQ ID NO:50), ORF5899 (SEQ ID NO:52), ORF6325 (SEQ ID NO:54), ORF7567c (SEQ ID NO:56), ORF7180 (SEQ ID NO:58), ORF7501 (SEQ ID NO:60), ORF7584 (SEQ ID NO:62), ORF8208c (SEQ ID NO:64), ORF8109 (SEQ ID NO:66), ORF9005c (SEQ ID NO:68), ORF8222 (SEQ ID NO:70), ORF8755c (SEQ ID NO:72), ORF9431c (SEQ ID NO:74), ORF9158 (SEQ ID NO:76), ORF10125c (SEQ ID NO:78), ORF9770 (SEQ ID NO:80), ORF9991 (SEQ ID NO:82), ORF10765c (SEQ ID NO:84), ORF10475 (SEQ ID NO:86), ORF11095c (SEQ ID NO:88), ORF11264 (SEQ ID NO:90), ORF11738 (SEQ ID NO:92), ORF12348c (SEQ ID NO:94), ORF12314c (SEQ ID NO:96), ORF13156c (SEQ ID NO:98), ORF12795 (SEQ ID NO:100), ORF13755c (SEQ ID NO:210), ORF13795c (SEQ ID NO:212), ORF14727c (SEQ ID NO:214), ORF13779 (SEQ ID NO:216), ORF14293c (SEQ ID NO:218), ORF14155 (SEQ ID NO:220), ORF14360 (SEQ ID NO:222), ORF15342c (SEQ ID NO:224), ORF15260c (SEQ ID NO:226), ORF14991 (SEQ ID NO:228), ORF15590c (SEQ ID NO:230), ORF15675c (SEQ ID NO:232), ORF16405 (SEQ ID NO:234), ORF16925 (SEQ ID NO:236), ORF17793c (SEQ ID NO:238), ORF18548c (SEQ ID NO:240), ORF17875 (SEQ ID NO:242), ORF18479 (SEQ ID

NO:244), ORF19027c (SEQ ID NO:246), ORF19305 (SEQ ID NO:248), ORF19519 (SEQ ID NO:250), ORF19544 (SEQ ID NO:252), ORF20008 (SEQ ID NO:254), ORF20623c (SEQ ID NO:256), ORF21210c (SEQ ID NO:258), ORF21493c (SEQ ID NO:260), ORF21333 (SEQ ID NO:262), ORF22074c (SEQ ID NO:264), ORF21421 (SEQ ID NO:266), ORF22608c (SEQ ID NO:268), ORF22626 (SEQ ID NO:270), ORF23228 (SEQ ID NO:272),ORF23367 (SEQ ID NO:274), ORF25103c (SEQ ID NO:276), ORF23556 (SEQ ID NO:278), ORF26191c (SEQ ID NO:280), ORF23751 (SEQ ID NO:282), ORF24222 (SEQ ID NO:284), ORF24368 (SEQ ID NO:286), ORF24888c (SEQ ID NO:288), ORF25398c (SEQ ID NO:290), ORF25892c (SEQ ID NO:292), ORF25110 (SEQ ID NO:294), ORF25510 (SEQ ID NO:296), ORF26762c (SEQ ID NO:298), ORF26257 (SEQ ID NO:300), ORF26844c (SEQ ID NO:302), ORF26486 (SEQ ID NO:304), ORF26857c (SEQ ID NO:306), ORF27314c (SEQ ID NO:308), ORF27730c (SEQ ID NO:310), ORF26983 (SEQ ID NO:312), ORF28068c (SEQ ID NO:314), ORF27522 (SEQ ID NO:316), ORF28033c (SEQ ID NO:318), ORF29701c (SEQ ID NO:320), ORF28118 (SEQ ID NO:322), ORF28129 (SEQ ID NO:324), ORF29709c (SEQ ID NO:326), ORF29189 (SEQ ID NO:328), ORF29382 (SEQ ID NO:330), ORF30590c (SEQ ID NO:332), ORF29729 (SEQ ID NO:334), ORF30221 (SEQ ID NO:336), ORF30736c (SEQ ID NO:338), ORF30539 (SEQ ID NO:340), ORF31247c (SEQ ID NO:342), ORF31539c (SEQ ID NO:346), ORF31222 (SEQ ID NO:348), ORF31266 (SEQ ID NO:350), ORF31661c (SEQ ID NO:352), ORF32061c (SEQ ID NO:354), ORF32072c (SEQ ID NO:356), ORF31784 (SEQ ID NO:358), ORF32568c (SEQ ID NO:360), ORF33157c (SEQ ID NO:362), ORF32539 (SEQ ID NO:364), ORF33705c (SEQ ID NO:366), ORF32832 (SEQ ID NO:368), ORF33547c (SEQ ID NO:370), ORF33205 (SEQ ID NO:372), ORF33512 (SEQ ID NO:374), ORF33771 (SEQ ID NO:376), ORF34385c (SEQ ID NO:378), ORF33988 (SEQ ID NO:380), ORF34274 (SEQ ID NO:382), ORF34726c (SEQ ID NO:384), ORF34916 (SEQ ID NO:386), ORF35464c (SEQ ID NO:388), ORF35289 (SEQ ID

NO:390), ORF35410 (SEQ ID NO:392), ORF35907c (SEQ ID NO:394),ORF35534 (SEQ ID NO:396), ORF35930 (SEQ ID NO:398), ORF36246 (SEQ ID NO:400), ORF26640c (SEQ ID NO:402), ORF36739 (SEQ ID NO:404), ORF37932c (SEQ ID NO:406), ORF38640c (SEQ ID NO:408), ORF39309c (SEQ ID NO:410), ORF38768 (SEQ ID NO:412), ORF40047c (SEQ ID NO:414), ORF40560c (SEQ ID NO:416), ORF40238 (SEQ ID NO:418), ORF40329 (SEQ ID NO:420), ORF40709c (SEQ ID NO:422), ORF40507 (SEQ ID NO:424), ORF41275c (SEQ ID NO:426), ORF42234c (SEQ ID NO:428), ORF41764c (SEQ ID NO:430), ORF41284 (SEQ ID NO:432), ORF41598 (SEQ ID NO:434), ORF42172c (SEQ ID NO:436), and ORF42233c (SEQ ID NO:438).

Fig. 4 shows the deduced amino acid sequences for ORF2 (SEQ ID NO:3), ORF3 (SEQ ID NO:5), ORF602c (SEQ ID NO:7), ORF214 (SEQ ID NO:9),ORF1242c (SEQ ID NO:11), ORF594 (SEQ ID NO:13), ORF1040 (SEQ ID NO:15), ORF1640c (SEQ ID NO:17), ORF2228c (SEQ ID NO:19), ORF2068c (SEQ ID NO:21), ORF1997 (SEQ ID NO:23), ORF2558c (SEQ ID NO:25), ORF2929c (SEQ ID NO:27), ORF3965c (SEQ ID NO:29), ORF3218 (SEQ ID NO:31), ORF3568 (SEQ ID NO:33), ORF4506c (SEQ ID NO:35), ORF3973 (SEQ ID NO:37), ORF4271 (SEQ ID NO:39), ORF4698 (SEQ ID NO:41), ORF5028 (SEQ ID NO:43), ORF5080 (SEQ ID NO:45), ORF6479c (SEQ ID NO:47), ORF5496 (SEQ ID NO:49), ORF5840 (SEQ ID NO:51), ORF5899 (SEQ ID NO:53), ORF6325 (SEQ ID NO:55), ORF7567c (SEQ ID NO:57), ORF7180 (SEQ ID NO:59), ORF7501 (SEQ ID NO:61), ORF7584 (SEQ ID NO:63), ORF8208c (SEQ ID NO:65), ORF8109 (SEQ ID NO:67), ORF9005c (SEQ ID NO:69), ORF8222 (SEQ ID NO:71), ORF8755c (SEQ ID NO:73), ORF9431c (SEQ ID NO:75), ORF9158 (SEQ ID NO:77), ORF10125c (SEQ ID NO:79), ORF9770 (SEQ ID NO:81), ORF9991 (SEQ ID NO:83), ORF10765c (SEQ ID NO:85), ORF10475 (SEQ ID NO:87), ORF11095c (SEQ ID NO:89), ORF11264 (SEQ ID NO:91), ORF11738 (SEQ ID NO:93), ORF12348c (SEQ ID NO:95), ORF12314c (SEQ ID NO:97), ORF13156c (SEQ

ID NO:99), ORF12795 (SEQ ID NO:101), ORF13755c (SEQ ID NO:211), ORF13795c (SEQ ID NO:213), ORF14727c (SEQ ID NO:215), ORF13779 (SEQ ID NO:217), ORF14293c (SEQ ID NO:219), ORF14155 (SEQ ID NO:221), ORF14360 (SEQ ID NO:223), ORF15342c (SEQ ID NO:225), ORF15260c (SEQ ID NO:227), ORF14991 (SEQ ID NO:229), ORF15590c (SEQ ID NO:231), ORF15675c (SEQ ID NO:233), ORF16405 (SEQ ID NO:235), ORF16925 (SEQ ID NO:237), ORF17793c (SEQ ID NO:239), ORF18548c (SEQ ID NO:241), ORF17875 (SEQ ID NO:243), ORF18479 (SEQ ID NO:245), ORF19027c (SEQ ID NO:247), ORF19305 (SEQ ID NO:249), ORF19519 (SEQ ID NO:251), ORF19544 (SEQ ID NO:253), ORF20008 (SEQ ID NO:255), ORF20623c (SEQ ID NO:257), ORF21210c (SEQ ID NO:259), ORF21493c (SEQ ID NO:261), ORF21333 (SEQ ID NO:263), ORF22074c (SEQ ID NO:265), ORF21421 (SEQ ID NO:267), ORF22608c (SEQ ID NO:269), ORF22626 (SEQ ID NO:271), ORF23228 (SEQ ID NO:273), ORF23367 (SEQ ID NO:275), ORF25103c (SEQ ID NO:277), ORF23556 (SEQ ID NO:279), ORF26191c (SEQ ID NO:281), ORF23751 (SEQ ID NO:283), ORF24222 (SEQ ID NO:285), ORF24368 (SEQ ID NO:287), ORF24888c (SEQ ID NO:289), ORF25398c (SEQ ID NO:291), ORF25892c (SEQ ID NO:293), ORF25110 (SEQ ID NO:295), ORF25510 (SEQ ID NO:297), ORF26762c (SEQ ID NO:299), ORF26257 (SEQ ID NO:301), ORF26844c (SEQ ID NO:303), ORF26486 (SEQ ID NO:305), ORF26857c (SEQ ID NO:307), ORF27314c (SEQ ID NO:309), ORF27730c (SEQ ID NO:311), ORF26983 (SEQ ID NO:313), ORF28068c (SEQ ID NO:315), ORF27522 (SEQ ID NO:317), ORF28033c (SEQ ID NO:319), ORF29701c (SEQ ID NO:321), ORF28118 (SEQ ID NO:323), ORF28129 (SEQ ID NO:325), ORF29709c (SEQ ID NO:327), ORF29189 (SEQ ID NO:329), ORF29382 (SEQ ID NO:331), ORF30590c (SEQ ID NO:333), ORF29729 (SEQ ID NO:335), ORF30221 (SEQ ID NO:337), ORF30736c (SEQ ID NO:339), ORF30539 (SEQ ID NO:341), ORF31247c (SEQ ID NO:343), ORF30963c (SEQ ID NO:345), ORF31539c (SEQ ID NO:347), ORF31222 (SEQ ID NO:349), ORF31266 (SEQ ID

NO:351), ORF31661c (SEQ ID NO:353), ORF32061c (SEQ ID NO:355), ORF32072c (SEQ ID NO:357), ORF31784 (SEQ ID NO:359), ORF32568c (SEQ ID NO:361), ORF33157c (SEQ ID NO:363), ORF32530 (SEQ ID NO:365), ORF33705c (SEQ ID NO:367), ORF32832 (SEQ ID NO:369), ORF33547c (SEQ ID NO:371), ORF33205 (SEQ ID NO:373), ORF33512 (SEQ ID NO:375), ORF33771 (SEQ ID NO:377), ORF34385c (SEQ ID NO:379), ORF33988 (SEQ ID NO:381), ORF34274 (SEQ ID NO:383), ORF34726c (SEQ ID NO:385), ORF34916 (SEQ ID NO:387), ORF35464c (SEQ ID NO:389), ORF35289 (SEQ ID NO:391), ORF35410 (SEQ ID NO:393), ORF35907c (SEQ ID NO:395), ORF35534 (SEQ ID NO:397), ORF35930 (SEQ ID NO:399), ORF36246 (SEQ ID NO:401), ORF26640c (SEQ ID NO:403), ORF36769 (SEQ ID NO:405), ORF37932c (SEQ ID NO:407), ORF38640c (SEQ ID NO:409), ORF39309c (SEQ ID NO:411), ORF38768 (SEQ ID NO:413), ORF40047c (SEQ ID NO:415), ORF40560c (SEQ ID NO:417), ORF40238 (SEQ ID NO:419), ORF40329 (SEQ ID NO:421), ORF40709c (SEQ ID NO:423), ORF40507 (SEQ ID NO:425), ORF41275c (SEQ ID NO:427), ORF42234c (SEQ ID NO:429), ORF41764c (SEQ ID NO:431), ORF41284 (SEQ ID NO:433), ORF41598 (SEQ ID NO:435), ORF42172c (SEQ ID NO:437), and ORF42233c (SEQ ID NO:439).

Fig. 5 shows the nucleotide sequence (SEQ ID NO:102) encoding a protein encoded by the 33A9 sequence.

Fig. 6A shows the deduced amino acid sequence (SEQ ID NO:103) a protein encoded by the 33A9 sequence.

Fig. 6B shows the nucleotide sequences of several ORFs1-10 (SEQ ID NOS:189, 190, 191, 192, 193, 194, 195, 196, 197, and 198) identified in the 33A9 sequence and their respective amino acid sequences (ORFs1-10; SEQ ID NOS:199, 200, 201, 202, 203, 204, 205, 206, 207, and 208).

Fig. 7 shows the physical map of the 34B12 EcoR1 fragment map identifying the positions of three ORFs: ORF1 (L-S), ORF2, and ORF 1S. The nucleotide sequence

corresponding to the *pho34B12* insertion (SEQ ID NO:104) containing ORF1 (L-S) (SEQ ID NOS:105 and 107), ORF2 (SEQ ID NOS:106 and 108), and ORF1-S(SEQ ID NOS:208 and 209) is also shown.

Fig. 8 shows the deduced amino acid sequence of ORF1(L-S) (SEQ ID NO:107) which is depicted in Fig. 7.

Fig. 9 shows the deduced amino acid sequence of ORF2 (SEQ ID NO:108) which is depicted in Fig. 7.

Fig. 10 shows the nucleotide sequence (SEQ ID NO:109) corresponding to the *36A4* insertion.

Fig. 11 shows the deduced amino acid sequence of the peptide (SEQ ID NO: 110) encoded by the 36A4 sequence. The predicted peptide encoded by the 36A4 sequence has homology to the *hrpM* gene of *Pseudomonas syringae* (Loubens, et al. *Mol. Microbiol.* 10: 329-340, 1993).

Fig. 12 shows the nucleotide sequence (SEQ ID NO:111) of contig 2507 identified using 36A4 nucleotide sequence.

Fig. 13 shows the nucleotide sequence (SEQ ID NO:112) corresponding to the *23A2* insertion.

Fig. 14A shows the deduced amino acid sequence of the peptide (SEQ ID NO: 113) encoded by the *23A2* sequence. The peptide predicted by the *23A2* sequence is homologous to a known protein in *Pseudomonas aeruginosa* (strain CD10): the *mexA* gene. This gene is part of an operon that also contains two other genes: *mexB* and *oprM* (Poole et al., *Mol. Microbiol.* 10: 529-544, 1993); GenBank submission: L11616.

Fig. 14B shows the nucleotide sequence (SEQ ID NO:148) and predicted partial amino acid sequences of PA14 *mexA and mexB* (SEQ ID NOS: 149 and 150, respectively).

Fig. 15 shows the nucleotide sequence (SEQ ID NO:114) of the PAO1 phenazine operon that was identified using the 3E8 sequence tag.

Fig. 16A shows the nucleotide sequence (SEQ ID NO:115) of the 3E8 sequence tag.

Fig. 16B shows the nucleotide sequences flanking the 3E8 sequence tag (SEQ ID NO:160).

Fig. 17 shows the deduced 3E8 PHZA amino acid sequence (SEQ ID NO: 116).

Fig. 18A shows the deduced 3E8 PHZB amino acid sequence (SEQ ID NO: 117).

Fig. 18B shows the deduced 3E8 PHZA partial amino acid sequence (SEQ ID NO:161).

Fig. 18C shows the deduced 3E8 PHZB partial amino acid sequence (SEQ ID NO:162).

Fig. 18D shows the deduced 3E8 PHZC partial amino acid sequence (SEQ ID NO:163).

Fig. 18E shows the nucleotide sequence (SEQ ID NO:164) and predicted partial amino acid sequence (SEQ ID NO:165) of PA14 *phzR*.

Fig. 19 shows the nucleotide sequence (SEQ ID NO:118) of the 34H4 sequence tag.

Fig. 20 shows the nucleotide sequence (SEQ ID NO:119) of the 33C7 sequence tag.

Fig. 21 shows the nucleotide sequence (SEQ ID NO:120) of the 25a12.3 sequence tag.

Fig. 22 shows the nucleotide sequence (SEQ ID NO:121) of the 8C12 sequence tag.

Fig. 23 shows the nucleotide sequence (SEQ ID NO:122) of the 2A8 sequence tag.

Fig. 24A shows the nucleotide sequences (SEQ ID NOS:123, 124, 125, 126,

127, and 128) of the 41A5, 50E12, 35A9, pho23, 16G12, and 25F1 Tn*phoA* sequence tags, respectively.

Fig. 24B shows the nucleotide sequence (SEQ ID NO:166) and predicted amino acid sequence (SEQ ID NO:167) of PA14 *pho15*.

Fig. 24C shows the nucleotide sequence (SEQ ID NO:168) of PA14 50E12 encoding $YgdP_{Pa}$ (SEQ ID NO:169) and $PtsP_{Pa}$ (SEQ ID NO:170).

Fig. 24D shows the nucleotide sequence (SEQ ID NO:171) of PA14 35A9, encoding $mtrR_{Pa}$ (SEQ ID NO:172).

Fig. 24E shows the nucleotide sequence (SEQ ID NO:173) of PA14 25F1 encoding ORFT (SEQ ID NO:174), ORFU (SEQ ID NO:175), and $DjlA_{Pa}$ (SEQ ID NO:176).

Fig. 25 shows the nucleotide sequence (SEQ ID NO:129) of the *phnA* and *phnB* genes of *Pseudomonas aeruginosa* of PAO1 and PA14, respectively.

Fig. 26 shows the deduced amino acid sequence (SEQ ID NO:130) of PHNA.

Fig. 27 shows the nucleotide sequence (SEQ ID NO:131) of the PA14 *degP* gene.

Fig. 28 shows the deduced amino acid sequence (SEQ ID NO:132) of the PA14 *degP* gene.

Fig. 29 shows the nucleotide sequence (SEQ ID NO:133) of the *algD* gene of *Pseudomonas aeruginosa* strain 8830.

Fig. 30 shows the deduced amino acid sequence (SEQ ID NO:134) of the *algD* gene of *Pseudomonas aeruginosa* strain 8830.

Fig. 31 shows the nucleotide sequence (SEQ ID NO:135) of the 1126 contig identified using 25A12.

Fig. 32 shows the physical map of the 1344 (SEQ ID NO:136) contig identified using 33C7 which illustrates three identified ORFs: ORFA (SEQ ID NO:440), ORFB (SEQ ID NO:441), and ORFC (SEQ ID NO:442). The amino acid sequences of

ORFA (SEQ ID NO:443), ORFB (SEQ ID NO:444), and ORFC (SEQ ID NO:445) encoded by their respective ORF is also shown.

Fig. 33 shows the nucleotide sequence (SEQ ID NO:137) of the 1G2 sequence tag.

Figs. 34A-D are graphs showing the complementation of the worm pathogenicity phenotype of 4 Tn*phoA* mutants using the *C. elegans* slow-killing assay.

Fig. 34A is a graph showing that the nonpathogenic phenotype of mutant *12A1* (open diamonds) could be fully complemented to the wild-type PA14 levels (filled squares) by the *lasR* gene from PAO1 under the control of the constitutive *lacZ* promoter *in trans* in strain *12A1* (pKDT17) (open circles). The reconstructed *lasR* mutant, PA14 *lasR-G* (open squares) is as nonpathogenic as *12A1* (open diamonds). Results from an experiment using one-day-old adults is shown.

Fig. 34B is a graph showing the complementation of the delayed-killing phenotype of *pho15*. Strains *pho15*(pEcdsbA) (open diamonds) and *pho15*(pPAdsbA), carry the *dsbA* gene from *E. coli* and *P. aeruginosa*, respectively, *in trans* under the control of the constitutive *lacZ* promoter.

Fig. 34C is a graph showing that the delayed killing phenotype of *25F1* was only partially restored by strains *25F1*(pORF338) and *25F1*(p3-ORFs) carrying plasmids containing *orf338* and *orf338-orf224-djlA*$_{Pa}$, respectively.

Fig. 34D is a graph showing the complementation of *50E12* by the *orf159-ptsP*$_{Pa}$ operon. Strain *50e12*(pUCP18), like mutant *12A1*, does not kill worms even after 63 hours. Both strains *50E12*(pMT205-lac) and *50E12*(pMT206-nat), expressing the putative *orf159-ptsP*$_{Pa}$ operon were able to kill *C. elegans*. In *50E12*(pMT205-lac), transcription of *orf159-ptsP*$_{Pa}$ is under the control of the constitutive *lacZ* promoter, whereas in *50E12*(pMT206-nat), the operon is controlled by its native promoter. Each data point represents means ± SD of 3-4 replicates. Unless indicated otherwise, synchronized L4 worms were used in the experiments. At least two independent

experiments were performed for each complementation analysis.

Fig. 35A is a schematic illustration showing the anthranilate synthase complex that is encoded by the *phnA* and *phnB* genes which catalyzes the conversion of chorismate to anthranilate. Antranilate serves as a precursor for pyocyanin production in *P. aeruginosa*, strain PAO1 (Essar et al., *J. Bacteriol.* 172: 884-900, 1990). The double arrows indicate the involvement of multiple, undefined steps, leading from the conversion of anthranilate to pyocyanin.

Fig. 35B is a schematic illustration showing the generation of the *ΔphnAphnB* mutant by an in-frame deletion of 1602 bp within the *phnA* and *phnB* genes.

Fig. 35C is a graph showing the effect of the *ΔphnAphnB* mutant on fast killing in *C. elegans*. Fast- killing assays were conducted using the wild type PA14 strain, the Tn*phoA* mutant *3E8* or the *ΔphnAphnB* strain. Worm mortality was monitored 3 hours after initial exposure to the bacteria and the defect in fast killing seen with *ΔphnAphnB* strain was comparable to that of another phenazine mutant, *3E8*.

Virulence Factor Identification and Characterization

As described herein, plants were used as an *in vivo* pathogenesis model for the identification of virulence factors of the human opportunistic pathogen *Pseudomonas aeruginosa*. Nine out of nine *TnphoA* mutant derivatives of *P. aeruginosa* strain UCBPP-PA14 that were identified in a plant leaf assay for less pathogenic mutants also exhibited significantly reduced pathogenicity in a mouse burn assay, suggesting that *P. aeruginosa* utilized many common strategies to infect both hosts. Seven of these nine mutants contained *TnphoA* insertions in previously unknown genes. These results demonstrated that an alternative non-vertebrate host of a human bacterial pathogen could be used in an *in vivo* high throughput screen to identify novel bacterial virulence factors involved in mammalian pathogenesis. These experimental examples are intended to illustrate, not limit, the scope of the claimed invention.

These experiments were carried out using the following techniques.

Strains, Growth Conditions and Plasmids. *P. aeruginosa* strain UCBPP-PA14 is a human clinical isolate that was used in these experiments for the identification of novel virulence-related genes (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995), and *P. aeruginosa* strains PAK (Ishimoto and Lory, *Proc. Natl. Acad. Sci. USA* 86:1954-1957, 1989) and PAO1 (Holloway et al., *Microbiol. Rev.* 43:73-102, 1979) have been studied extensively in many laboratories. Luria Bertani broth and agar were used for the growth of *P. aeruginosa* and *Escherichia coli* strains at 37°C. Minimal medium (M9) was also used for the growth of *P. aeruginosa.*

Transposon Mutagenesis. Transposon-mediated mutagenesis of UCBPP-PA14 was performed using Tn*phoA* carried on the suicide plasmid pRT731 in *E. coli* strain SM10 λpir (Taylor et al., *J. Bacteriol.* 171:1870-1878, 1989). Donor and recipient cells grown in this medium were plated together on Luria Bertani agar plates and incubated at 37°C for eight to ten hours and subsequently plated on Luria Bertani plates containing rifampicin (100 μg/ml) (to select against the *E. coli* donor cells) and kanamycin (200 μg/ml) (to select for Tn*phoA* containing *P. aeruginosa* cells). Colonies which grew on the rifampicin and kanamycin media were replicated to Luria Bertani containing ampicillin (300 μg/ml); ampicillin resistant colonies indicated pRT731 integration into the UCBPP-PA14 genome and were discarded.

Alkaline Phosphatase Activity. Two thousand five hundred (2,500) prototrophic UCBPP-PA14 Tn*phoA* mutants were screened on peptone glucose agar plates (Ostroff et al., *J. Bacteriol.* 172:5915-5923, 1990) containing 40 μg/ml 5-bromo-4-chloro-3-indoly phosphate (XP). Peptone medium was selected because it suppressed the production of the endogenous blue-green pigment pyocyanin and the

fluorescent yellow pigment pyoverdin, permitting visualization of the blue color that resulted from dephosphorylation of XP by periplasmic alkaline phosphatase generated by $PhoA^+$ mutants.

Growth Conditions and Mutant Isolation Strategy. *P. aeruginosa* strains that were grown to saturation in L-broth at 37°C were washed in 10 mM $MgSO_4$, resuspended at an optical density of 0.2 (OD_{600} = 0.2) in 10 mM $MgSO_4$ and diluted 1:100 and 1:1000 (corresponding to a bacterial density of approximately 10^6 and 10^5 cfu/ml, respectively). Approximately 10 ml of the diluted cells were inoculated with a Pipetman into stems of approximately twelve-week old lettuce plants (variety Romain or Great lake) grown in MetroMix potting soil in a greenhouse (26°C). The stems were washed with 0.1% bleach and placed on 15 cm diameter petri dishes containing one Whatman filter (Whatman #1) that was impregnated with 10 mM $MgSO_4$. The midrib of each lettuce leaf was inoculated with three different Tn*phoA*-generated *P. aeruginosa* mutants to be tested and the wild type UCBPP-PA14 strain as a control. The plates were kept in a growth chamber during the course of the experiment at 28-30°C and 90-100% relative humidity. Symptoms were monitored daily for five days.

In the *Arabidopsis* leaf infiltration model, *P. aeruginosa* strains grown and washed as above were diluted 1:100 in 10 mM $MgSO_4$ (corresponding to a bacterial density of 10^3 /cm^2 leaf disk area) and were injected into leaves of six-week old *Arabidopsis* plants as described for infiltration of *Pseudomonas syringae* (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995; Dong et al., *Plant Cell* 3:61-72, 1991). Incubation conditions and monitoring of symptoms were the same as in the lettuce experiments. Leaf intercellular fluid containing bacteria was harvested, and bacterial counts were determined as described (Rahme et al., *Science* 268:1899-1902, 1995; Dong et al., *Plant Cell* 3:61-72, 1991). Four different samples

were taken using two leaf discs per sample. Control plants inoculated with 10 mM $MgSO_4$ showed no symptom developement.

Mice Mortality Studies. A 5% total surface area burn was fashioned on the outstreached abdominal skin of six-week-old male AKR/J mice (Jackson Laboratories) weighing between 25 and 30 gm as previously described (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995; Stevens, *J. Burn Care Rehabil.* 15:232-235, 1994). Immediately following the burn, mice were injected with 5×10^3 or 5×10^5 *P. aeruginosa* cells, and the number of animals that died of sepsis was monitored each day for ten days. Animal study protocols were reviewed and approved by the subcommittee on Animal Studies of the Massachusetts General Hospital. Statistical significance for mortality data was determined using a χ^2 test with Yates' correction or Fisher's exact test. Differences between groups were considered statistically significant at $P \leq 0.05$.

DNA Manipulation, Molecular Cloning, and Sequence Analysis of TnphoA Mutants. *P. aeruginosa* chromosomal DNA was isolated by phenol extraction (Strom and Lory, *J. Bacteriol.* 165:367-372, 1986), and DNA blotting and hybridization studies were performed as described in Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley, New York, 1996).

The oligonucleotides 5'- AATATCGCCCTGAGCAGC- 3' (LGR1) (SEQ ID NO:138) and 5' -AATACACTCACTATGCGCTG- 3' (LGR2) (SEQ ID NO:139) corresponded to sequences on opposite strands at the 5'- end of Tn*phoA*. The oligonucleotides 5'-CCATCTCATCAGAGGGTA-3' (LGR3) (SEQ ID NO:140) and 5'-CGTTACCATGTTAGGAGGTC-3' (LGR4) (SEQ ID NO:141) corresponded to sequences on opposite strands at the of the 3'- end of Tn*phoA*. LGR1 + LGR2 or LGR3 + LGR4 were used to amplify by inverse PCR (IPCR) DNA sequences adjacent to the sites

of Tn*phoA* insertion as described (Ochman et al., 1993, *A Guide to Methods and Applications*, eds., Innis, M.A., States, D.J., 1990). Amplified DNA fragments ranging in size from 350 to 650 base pairs were cloned into pBlueScript SK+/- by filling in the ends of the IPCR products prior to subcloning into the EcoRV site of pBlueScript SK+/-. To determine the sequence of IPCR-amplified products, double-stranded DNA sequencing was performed using the Sequenase 2.0 kit (U. S. Biochemical, Inc.). Sequences obtained were compared to the non-redundant peptide sequence databases at the National Center for Biotechnology Information (NCBI) using the BLASTX program (Gish and States, *Nat. Genet.* 3:266-272, 1993).

Isolation and DNA Manipulation of the Wild Type Clone Containing the Gene Corresponding to the pho34B12 Mutation from the UCBPP-PA14 Genomic Library. The IPCR product that was generated from UCBPP-PA14 Tn*phoA* mutant *pho34B12* mutant was labeled using a random primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN) and used to probe a genomic library of UCBPP-PA14 chromosomal DNA in pJSR1 (Rahme et al., *Science* 268:1899-1902, 1995) for a clone containing the gene corresponding to the *pho34B12* mutation. A 3.7 kb *EcoRI* fragment, identified in cosmid clone pLGR34B12 which corresponded to the pho34B12 mutation, was subcloned into *EcoRI* site of pRR54 (Roberts et al., *J. Bacteriol.* 172:6204-6216, 1990) after filling-in the ends of both vector and fragment to construct pLGRE34B12. The same fragment (made blunt ended) was subcloned into the *SmaI* site of pCVD (Donnenberg and Kaper, *Infect. Immun.* 59:4310-4317, 1991) to construct pLGR34. pLGR34 was used to replace the mutated pho34B12 gene with a wild-type copy as described (Donnenberg and Kaper, *Infect. Immun.* 59:4310-4317, 1991). The 3.7 kb EcoRI fragment was also subcloned into the EcoRI site of pBlueScript SK+/- to construct pBSR34B12 and used for DNA sequence analysis.

A 1,659 base pair sequence corresponding to the *pho34B12* insertion that contains two overlapping open reading frames (ORF1 and ORF2) on opposing strands

was submitted to GenBank and was assigned Accession number AF031571. ORF1 is 1,148 bp (nucleotides 361 to 1509) and ORF2 is 1,022 bp (nucleotides 1458 to 436). The overlap of the two ORFs is from nucleotide 436 to 1458. ORF1 contains a second putative translational start site at nucleotide 751 corresponding to a coding region of 758 bp. The oligonucleotide primers 5'-CGCATCGTCGAAACGCTGGCGGCC-3' (SEQ ID NO:142) and 5'-GCCGATGGCGAGATCATGGCGATG-3' (SEQ ID NO:143) were used to amplify a 1100 bp fragment from pBSR34B12 containing ORF1. Because of the two putative initiation sites present in ORF1, the oligonucleotide primers 5'-TGCGCAACGATACGCCGTTGCCGACGATC-3' (SEQ ID NO:144) and 5'-GATTCCACCTTCGCAGCGCAGCCC-3' (Reg3) (SEQ ID NO:145) were also used to amplify a 1659 bp from pBSR34B12 containing ORF1. The oligonucleotide primers 5'-GATTCCACCTTCGCAGCGCAGCCC-3' (SEQ ID NO:146) and 5'-GCCGATGGCGAGATCATGGCGATG-3' (SEQ ID NO:147) were used to amplify a 1302 bp fragment from pBSR34B12 containing ORF2. All primer combinations were designed to contain the putative upstream regulatory elements of each ORF. The PCR products obtained (1100, 1659, and 1302 bp) were cloned into pCR2.1 (Invitrogen Inc.) to construct pLE15, pLE1, and pLE2, respectively. All three PCR products were subcloned into pRR54 to construct pRRLE15, pRRLE1, and pRRLE2, respectively.

Enzymatic Activities of Tn*phoA* Mutants. *P. aeruginosa* strains grown for eighteen hours in LB medium were used for assays of enzymatic activities. Proteolytic and elastolytic activities were determined as described previously (Toder et al., *Mol. Microbiol.* 5:2003-2010, 1991). Quantitation of pyocyanin was determined as described (Essar et al., *J. Bact.* 172:884-900, 1990). Hemolytic activity was detected following incubation on plates containing Trypticase soy agar (BBL) supplemented with 5% Sheep red blood cells (Ostroff and Vasil, *J. Bacteriol.* 169:4957-4601, 1987).

Generation of a Non-Polar *GacA* Mutation. A non-polar *gacA* mutation in UC BPP-PA14 was constructed by cloning a 3.5 kb PstI fragment containing the *gacA* gene

from cosmid pLGR43 (Rahme et al., *Science* 268:1899-1902, 1995) into the unique *BamHI* restriction site in the suicide vector pEGBR (Akerley et al., *Cell* 80:611-620, 1995) using *BamHI* linkers. A 950 bp *EcoRI-HincII* Klenow end-filled fragment containing the kanamycin resistance gene cassette from pUC18K (Menard et al., *J. Bacteriol.* 175:5899-5906, 1993) was then cloned into the unique *BamH1* restriction site (made blunt ended) in *gacA*, such that transcription was maintained and translation of the downstream portion of *gacA* was reinitiated at the 3' end of the kanamycin cassette. The resultant construct, SW 7-4, containing the kanamycin gene cassette within the *gacA* gene and in the orientation of its transcription, was used to marker-exchange by homologous recombination the disrupted *gacA* gene into the wild-type UCBPP-PA14 genome.

Isolation and Characterization of *P. aeruginosa* Virulence Factors. Using the procedures described above, the *P. aeruginosa* UCBPP-PA14 genome was mutagenized with transposon Tn*phoA*, and 2,500 prototrophic mutants were screened for impaired pathogenicity in the lettuce stem assay. This lettuce assay allowed for the testing of several mutants on a single lettuce stem. Interestingly, we found that lettuce was not only susceptible to infection by UCBPP-PA14 but also was susceptible to the well characterized *P. aeruginosa* strains PAK (Ishimoto and Lory, *Proc. Natl. Acad. Sci USA* 86:1954-1957, 1989) and PAO1 (Holloway et al., *Microbiol. Rev.* 43:73, 1979). Both of these latter strains proliferated in lettuce leaves and elicited disease symptoms similar to those elicited by UCBPP-PA14, characterized by water soaking followed by soft rot four to five days post-infection. In later stages of infection, all three *P. aeruginosa* strains invaded the entire midrib of a lettuce leaf resulting in complete maceration and collapse of the tissue.

As summarized in Table 1, we identified nine Tn*phoA*-generated mutants of UCBPP-PA14 among the 2,500 prototrophs screened that elicited either null, weak, or moderate rotting symptoms on lettuce stems compared to the wild-type strain.

Table 1

Strain	Growth in *Arabidopsis* leaves[a]	Symptoms Elicited in *Arabidopsis*[b]	% Mouse Mortality[c]		Gene Identity
			5×10^3	5×10^5	
PA14	5.5×10^7	severe	53	100	
33C7	8.3×10^4	none	0	0	unknown[d]
1D7	7.5×10^5	weak	0	50	*gacA*
25A12	1.7×10^6	weak	11	87	unknown
33A9	5.1×10^6	moderate	0	0	unknown
25F1	1.5×10^4	moderate	0	20	unknown
34H4	3.8×10^6	moderate	0	33	unknown
pho34B12	4.0×10^6	moderate	0	56	unknown
pho15	3.9×10^4	moderate	0	62	*dsbA*
16G12	2.3×10^5	moderate	20	100	unknown

[a]Four different samples were taken using two leaf discs/sample. Control plants inoculated with 10 mM $MgSO_4$ showed no symptoms during the course of the experiments. Three independent experiments gave similar results.

[b]Symptoms observed four to five days after infection. None, no symptoms; chlorosis, chlorosis circumscribing the inoculation site; weak, localized water-soaking and chlorosis of tissue circumscribing the inoculation site; moderate, moderate water-soaking and chlorosis with most of the tissue softened around the inoculation site; severe, severe soft-rotting of the entire leaf characterized by a water-soaked reaction zone and chlorosis around the inoculation site at two to three days post-infection.

[c]All animal experiments were conducted at least twice using 8-10 animals/experiment. Independent experiments showed similar percentage mortality rates. Mice were injected with ~ 5×10^3 or 5×10^5 cells.

[d]BLASTX analysis yielded no encoded proteins with significant homology.

Severe maceration of the leaf was not observed with any of the mutants. DNA blot analysis showed that each of the nine mutants contained a single Tn*phoA* insertion, using as a probe a 1542 base pair *BglI-BamHI* fragment containing the kanamycin resistance conferring gene of Tn*phoA* (Taylor et al., *J. Bact.* 171:1870-1878, 1989). Two of the nine UCBPP-PA14 Tn*phoA* mutants, *pho34B1,* and *pho15*, expressed alkaline phosphatase activity suggesting that the genes containing these Tn*phoA* insertions encoded membrane-spanning or secreted proteins (Taylor et al., *J. Bact.* 171:1870-1878, 1989; Manoil and Beckwith, *Proc. Natl. Acad. Sci USA* 82:5117, 1985).

The nine Tn*phoA* mutants were further tested by measuring their growth rate over the course of four days in *Arabidopsis* leaves as a quantitative measure of pathogenicity (Rahme et al., *Science* 268:1899-1902, 1995; Dong et al., *Plant Cell* 3:61-72, 1991). Although none of the mutants showed any significant differences in their growth rates as compared to the wild-type strain in both rich and minimal media, the growth rate over time of all nine mutants in *Arabidopsis* leaves was lower than the wild-type strain. Table 1 lists the maximal levels of growth reached by each mutant at the fourth day post-infection. In the case of all nine mutants, less severe symptom development reflected reduced bacterial counts in leaves. All of the mutants except *33C7* elicited either weak or moderate rot and water soaking symptoms with varying amounts of chlorosis (yellowing) (Table 1). Interestingly, however, as summarized in Table 1, the levels of proliferation of the individual mutants did not directly correlate with the severity of symptoms that they elicited. For example, even though mutant *25A12* (Fig. 21) grew to similar levels as mutants *33A9* (Figs. 5 and 6A-B), *pho34B12* (Figs. 7, 8, and 9), and *34H4* (Fig. 19), and only ten-fold less than wild-type UCBPP-PA14, mutant *25A12* elicited very weak symptoms. Similarly, mutants *33C7* (Fig. 20), *pho15* (Fig. 24B), and *25F1* (Fig. 24A) all reached similar maximal levels of growth (approximately 10^3-fold less than the growth of the wild type); however, only mutant *33C7* failed to cause any disease symptoms (Table 1). The differences observed in the degree of symptoms and proliferation levels among the ten mutants suggested that these mutants likely carried insertions in genes that are involved in various stages of the plant infectious process.

The pathogenicity of each of the nine Tn*phoA*-generated mutants that were less pathogenic in the plant leaf assay was measured in a full-thickness skin thermal burn mouse model (Rahme et al., *Science* 268:1899-1902, 1995; Stevens et al., *J. of Burn Care and Rehabil.* 15:232-235, 1994). As shown in Table 1, all nine mutants were significantly different from the wild-type with a $P \leq 0.05$ at both doses except for *25A12* and *16G12* (Fig. 24A), which were not significantly different from wild-type at the higher

dose of 5 x 10^5 cells. In addition to the data shown in Table 1, mutant *33A9* also caused no mortality even at a higher dose of 5 x 10^6.

We used DNA blot analysis and DNA sequence analysis to determine whether Tn*phoA* in the nine less pathogenic mutants had inserted in known genes. DNA blot analysis revealed that mutant *1D7* contained a Tn*phoA* insertion in the *gacA* gene (Laville et al., *Proc. Natl. Acad. Sci. USA* 89:1562-1566, 1992; Gaffney et al., *Mol. Plant-Microbe Interact.* 7:455-463, 1994) which we had shown previously to be an important pathogenicity factor for *P. aeruginosa* in both plants and animals (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995). For the other eight mutants we used the inverse polymerase chain reaction (IPCR) to generate amplified products corresponding to DNA sequences adjacent to the sites of the Tn*phoA* insertions (Ochman et al., *A Guide to Methods and Applications*, eds., Innis, M.A., States, D.J., 1990). The IPCR products were cloned and then subjected to DNA sequence analysis. Mutant *pho15* contained Tn*phoA* inserted into a *P. aeruginosa* gene (from strain PA01) previously deposited in GenBank (Accession # U84726) that shows a high degree similarity to the *Azotobacter vinelandii dsbA* gene, which encodes a periplasmic disulfide bond forming enzyme (Bardwell et al., *Cell* 67:581-589, 1991). Homologues of *dsbA* in the bacterial phytopathogen *Erwinia chrysanthemi* and in the human pathogens *Shigella flexneri* and *Vibrio cholera* are required for pathogenesis (Shevchik et al., *Mol. Microbiol* 16:745-753, 1995; Peek and Taylor, *Proc. Natl. Acad. Sci. USA* 89:6210-6214, 1992; Watarai et al., *Proc. Natl. Acad. Sci. USA* 92:4927-4931, 1995). Computer analysis using the program BLASTX showed that when the DNA sequences corresponding to the remaining seven Tn*phoA* insertions were translated in all possible reading frames, no significant similarities to any known genes were found (Table 1).

We performed a variety of biochemical tests to categorize the nine less

pathogenic UCBPP-PA14 mutants on the basis of whether they carried defects in previously described primary virulence factors and/or metabolic pathways. All mutants were assayed for protease, elastase, and phospholipase activities and for their ability to secrete the secondary metabolite pyocyanin (Toder et al., *Mol. Microbiol.* 5:2003-2010, 1991; Essar et al., *J. Bact.* 172:884-900, 1990; Ostroff and Vasil, *J. Bacteriol.* 169:45974601, 1987). Pyocyanin is a redox-active phenazine compound excreted by most clinical strains of *P. aeruginosa* that kills mammalian and bacterial cells through the generation of reactive oxygen intermediates and which has been implicated as a *P. aeruginosa* virulence factor (Hassett et al. *Infect. Immun.* 60:328-336, 1992; Kanthakumar et al., *Infect. Immun.* 61:2848-2853, 1993; Miller et al. *Infect. Immun.* 64:182, 1996). Mutants *33C7*, *33A9*, *34H4*, *25F1*, and *16G12* showed no defects in any of the biochemical assays used. Mutant *pho34B12* showed decreased hemolytic activity on blood agar plates, reduced elastase activity (~ 50%), and no detectable pyocyanin production. Mutant *pho15* showed only traces of elastase activity and a decrease in proteolytic activity (60-70%) compared to the wild-type. Mutant *25A12* showed a 50% decreased elastolytic activity. Finally, mutant *1D7* which contained an insertion in *gacA*, showed reduced levels of pyocyanin (~ 50%) as compared to the wild-type. In addition to mutant *1D7* a second independent *gacA*::Tn*phoA* mutant was identified from our plant screen, mutant *33D11*. This latter mutant also exhibited a similar reduction in pyocyanin production and reduced virulence in both plants and mice.

On the basis of the DNA sequence analysis and biochemical testing of the mutants, the genes targeted by the Tn*phoA* insertions in mutants *1D7* and *pho34B12* were chosen for further analysis. As discussed above, *1D7* contained an insertion in *gacA* which we had shown previously to encode a virulence factor in *P. aeruginosa* (Rahme et al., *Science* 268:1899-1902, 1995). Recently a *gacA*-like gene has also been shown to be an important virulence factor for *Salmonella typhimurium* (Johnston et al., *Mol. Microbiol.* 22:715, 1996). However, the two *gacA*::Tn*phoA* insertions (*1D7* and *33D11*),

the *gacA* insertion mutant that we constructed previously (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995), and an independently constructed *P. aeruginosa gacA* mutation that affects the production of several known virulence factors (Hassett et al., *Infect. Immun.* 60:328-336, 1992) all exert a polar effect on at least one gene, a homologue of the *E. coli uvrC* gene immediately downstream of *gacA* (Rahme et al., *Science* 268:1899-1902, 1995; Laville et al., *Proc. Natl. Acad. Sci USA* 89:1562-1566, 1992; Reimmann et al., *Mol. Microbiol.* 24:309-319, 1997). To provide definitive evidence that the loss of pathogenicity phenotypes of the *gacA* mutants described herein was due to the disruption of the *gacA* open reading frame per se rather than due to a polar effect on a gene downstream of *gacA*, we constructed a non-polar *gacA* mutation in UCBPP-PA14 using a DNA cassette encoding a gene that confers kanamycin resistance. Importantly, the non-polar *gacA* mutant exhibited the same diminished level of pathogenicity in the mouse assay (50% mortality) and in the Arabidopsis assay (growth to $3x10^5$ cfu/cm^2 after four days) as the *gacA*::Tn*phoA* mutant (*1D7*), but did not exhibit the extreme UV sensitivity of the polar *gacA* mutants. Like *1D7*, the non-polar *gacA* mutant also excreted lower levels of pyocyanin (50%) compared to the wild-type.

Mutant *pho34B12* was chosen for further analysis for the following reasons. First, the insertion in *pho34B12* was situated directly downstream of the *P. aeruginosa* pyocyanin biosynthetic genes *phnA* and *phnB* (Essar et al. *J. Bact.* 172:884-900, 1990), in a previously uncharacterized region of the *P. aeruginosa* genome. Second, the *pho34B12* insertion caused a pleiotropic phenotype that included reduced elastase and hemolytic activities, suggesting that the gene in which the *pho34B12* Tn*phoA* insertion was situated might encode a regulator of diverse pathogenicity factors.

To rule out the possibility that a secondary mutation in *pho34B12* was

responsible for the loss of pathogenicity phenotype rather than the Tn*phoA* insertion, we replaced the *pho34B12*::Tn*phoA* mutation by homologous recombination with the corresponding wild type gene. This resulted in restoration of the pathogenicity defect in both plants and animals as well as restoration of hemolytic and elastolytic activity and pyocyanin production to wild-type levels (Table 2, below).

Table 2[a]

Strain	Growth in *Arabidopsis* Leaves	Symptoms Elicited in *Arabidopsis*	% mouse mortality 5×10^5	% pyocyanin
PA14	5.5×10^7	severe	100	100
pho34B12	4.0×10^6	moderate	56	≤1
pho34B12 reconstructed to wild-type	3.9×10^7	severe	100	120
pho34B12 +*pLGRE34B12*	6.1×10^5	moderate	0	600
pho34B12 +pRRLE2	7.0×10^5	moderate	13	40
pho34B12 +*pRRLE1*	5.0×10^5	moderate	13	1,400
pho34B12 +*pRRLE15*	1.0×10^5	moderate	22	1,360

[a]See Table 1 for an explanation of table entries.

These results in Table 2 show that the Tn*phoA* insertion in *pho34B12* was the cause of the pleiotropic phenotype of this strain, including the loss of pathogenicity phenotype. The fact that no putative ORFs were present in the next 500 bp downstream of the stop codon following the *pho34B12*::Tn*phoA* insertion (see below) made it unlikely that Tn*phoA* exerted a polar effect on a downstream gene which was responsible for the phenotype of mutant *pho34B12*. Genetic complementation analysis of *pho34B12* with a plasmid (pLGRE34B12) containing a 3.7 kb insert which included *pho34B12* and part of

the *phnAB* region resulted in restoration of the elastase and hemolytic activities to wild-type levels and more than a ten-fold overproduction of pyocyanin (Table 2). However, the impaired pathogenicity phenotype of *pho34B12* in both *Arabidopsis* and mice was not complemented by pLGRE34B12 (Table 2), most likely due to the presence of multiple copies of the wild-type gene corresponding to *pho34B12*.

Further DNA sequence analysis showed that the region containing the *pho34B12* mutation encoded two almost completely overlapping open reading frames (ORFs) (ORF1 and ORF2) that were transcribed in opposite directions. Moreover, ORF1 had two potential methionine start codons (designated OFR1-S and ORF1-L). The predicted proteins encoded by ORF1-S and ORF1-L, which were transcribed in the same direction as the *phnA*, *phnB*, and *phoA* genes, contained a consensus motif that corresponded to a lipid attachment site found in a variety of prokaryotic membrane lipoproteins (Hayashi and Wu, *J. Bioenerg. Biomembr.* 22:451-471, 1990). These membrane lipoproteins are synthesized with a precursor signal peptide, providing an explanation for the Pho^+ phenotype of the *pho34B12* insertion (Hayashi and Wu, *J. Bioenerg. Biomembr.* 22:451-471, 1990). The predicted protein encoded by ORF2 contained an N-terminal 'helix-turn-helix' DNA-binding motif similar to the 'helix-turn-helix' motif found in the *LysR* family of transcriptional regulators (Henikoff et al., *Proc. Natl. Acad. Sci. USA* 85:6602-6606, 1988; Viale et al., *J. Bacteriol.*, 173:5224-5229, 1991). This class of proteins includes regulators involved in both mammalian and plant pathogenesis (Finlay and Falkow, *Microbiol. and Mol. Biol. Rev.* 61:136-169, 1997). The existence of two functional almost completely overlapping ORFs is unusual in bacterial genomes.

To determine which of the ORFs encoded in the *pho34B12* region were functional, additional complementation analysis was carried out using plasmids that contained PCR products corresponding to ORF1-S, ORF1-L, and ORF2 (Fig.7). The production of both pyocyanin and elastolytic activity was restored to 20-40% of wild type

levels by the plasmid synthesizing the protein encoded by ORF2 (pRRLE2). Similarly, the hemolytic ability of this complemented strain was partially restored. Complementation of *pho34B12* with plasmids pRRLE1 and PRRLE15, corresponding to ORF1-S and ORF1-L, respectively, also restored the hemolytic, pyocyanin, and elastolytic activities. Interestingly, however, the presence of plasmids pRRLE1 and pRRLE15 resulted in a 10-fold higher production of pyocyanin and a 2-fold higher level of elastase activity. Neither pRRLE1, pRRLE15, nor pRRLE2 complemented the loss of pathogenicity phenotypes of mutant *pho34B12* in either plants or animals (Table 2). Further characterization of this region including site directed mutagenesis will further elucidate which of the three ORFs is (are) required for pathogenicity in plants and animals.

The data presented above demonstrated that previously unknown *P. aeruginosa* virulence factors (genes) that play a significant role in mammalian pathogenesis can be readily identified by screening random *P. aeruginosa* mutants for ones that display attenuated pathogenic symptoms in plants. This is consistent with our previous study in which we demonstrated that at least three *P. aeruginosa* genes encode virulence factors involved in both plant and animal pathogenesis (Ausubel et al., *Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively; Rahme et al., *Science* 268:1899-1902, 1995). On the other hand, we did not expect to find that nine out of nine mutants that we isolated that were less virulent in plants would also be less virulent in mice. The simplest interpretation of this result is that *P. aeruginosa* pathogenesis in plants and animals utilizes a substantially overlapping set of genes which may be considered to be basic virulence genes. Another possible interpretation is that some of the identified genes may encode regulatory proteins (i.e., *pho34B12*), that control different effector molecules, a subset of which may be specific for either plants or animals. We also did not expect that

the majority of mutants that would be identified in this study (7 out of 9) would correspond to previously unknown genes. Using the Poisson distribution, a genome size for *P. aeruginosa* of 5.9 Mb and an average gene size of 1.1 kb, we calculated that the 2,500 mutants tested represents 25% of the total number that needs to be tested to give approximately 95% probability of testing each gene in the assay. Therefore, since our screen for *P. aeruginosa* virulence mutants is not nearly saturated, it is likely that many additional *P. aeruginosa* genes with important roles in pathogenicity await discovery.

Importantly, at least two of the previously known virulence factors (genes) identified in our model as being important in plant pathogenesis, are not only important virulence factors for *P. aeruginosa* in a mouse burn model, but have also been described as important virulence factors in other gram-negative pathogens. These latter pathogenicity factors (genes) include *dsbA*, and *gacA* (Shevchik et al. *Mol. Microbiol.* 16:745-753, 1995; Peek and Taylor, *Proc. Natl. Acad. Sci. USA* 89:6210-6214, 1992; Watarai et al., *Proc. Natl. Acad. Sci. USA* 92:4927-4931, 1995; Johnston, et al., *Mol. Microbiol.* 22:715, 1996). This makes it likely that many of the previously unknown factors identified in *P. aeruginosa* will be generally relevant for gram-negative pathogenesis.

Another important conclusion from this study is that the high throughput *in vivo* screening method that we have developed can lead to the identification of pathogenicity factors that do not correlate with obvious biochemical defects. Mutants *33C7, 33A9 , 34H4, 25F1,* and *16G12* exhibited no detectable defects in several known *P. aeruginosa* pathogenicity factors and, importantly, mutants *33C7* and *33A9* were among the most debilitated in the mouse model. Moreover, even though mutants *pho34B12* and *25A12* did exhibit diminished production of known virulence factors, the genes corresponding to these mutants have not been identified previously, most likely because the biochemical defects in these mutants cannot be readily identified efficiently in a simple high throughput screen. This attests to the sensitivity of our screen for loss of

pathogenicity phenotypes.

In the last few years, other high throughput screens for identifying bacterial pathogenicity factors have been described. The IVET (in vivo expression technology) identifies promoters that are specifically activated during pathogenesis (Wang et al., *Proc. Natl. Acad. Sci. USA.* 93:10434-10439, 1996; Mahan et al., *Science* 259:686-688, 1993), STM (signature-tagged transposon method) identifies genes that are required for survival in a host (Hensel, *Science* 268:400-403, 1995) and DFI (differential fluorescence induction) utilizes green fluorescent protein and fluorescence activated cell sorting to identify genes that are activated under specific conditions or in specific host cell types (Valdivia and Falkow, *Mol. Microbiol.* 22:367-378, 1996). These approaches are complimentary with the one that we have described in this application and each approach has advantages and disadvantages. One advantage of our screening procedure in a non-vertebrate host is that it directly measures pathogenicity whereas the IVET and DFI methods measure pathogenicity-associated gene expression. Unlike the STM procedure, which identifies genes whose function cannot be complemented in trans by the mixed population of bacterial mutants used for the inoculum, the present screen in a non-vertebrate involves testing each mutant clone separately.

Other Virulence Targets

The 33A9 nucleic acid sequence (Figs. 5 and 6A-B) was also identified in a cosmid clone designated BI48 (Fig. 1). This cosmid was sequenced in its entirety and its nucleic acid sequence is shown in Fig. 2. Using standard database analysis, the nucleotide sequences and deduced amino acid sequences of several additional open reading frames were identified (Figs. 3 and 4). A summary of this analysis is presented in Table 3. Like the sequences described above, any one of the sequences found in Figs. 3 and 4 can be used to screen for compounds (e.g., using the methods described herein) that reduce the virulence of a pathogen.

The sequence obtained from the pBI48 cosmid of strain PA14 revealed that 33A9 was located approximately 5 kb upstream of a pili gene cluster (Figure 1, Table 3). This cluster contains the *pilS/pilR* genes, known to be involved in the regulation of pili formation. Moreover, the analysis of the sequence upstream of 33A9 did not show any homology with previously identified sequences suggesting the possibility that the entire region surrounding 33A9 could define a pathogenicity island. Figs. 3 (orf 19544), Fig. 4 (orf 19544), 5, 6A, and 6B show the 33A9 nucleotide sequence, as well as the identified ORFs.

In addition, analysis of the sequence obtained from the pBI48 cosmid clone indicated the presence of a sequence located approximately 2 kb downstream of 33A9, which showed strong homology with tRNA sequences (ORF 22626, Figure 1). Because the analysis of the region located upstream of the tRNA sequence did not show any homology with sequences present in the database, and because tRNA sequences represent "hot spots" for DNA insertions, we hypothesized that the tRNA sequence represented the right boundary for the insertion of a pathogenicity island present in PA14. As seen in Figure 1 the size of the region that could represent the piece of foreign DNA that was inserted is approximately 25 kb. The identification of the boundary that is located upstream of the presumptive pathogenicity island will assist to establish the exact size of the inserted piece of DNA. Moreover, the analysis of the 33A9 region also indicated the presence of more than one sequence with homology at the protein level to integrases and transposases (ORF21421, ORF8109 respectively). Finally, our data showed that the 33A9 locus was present in several highly pathogenic *P. aeruginosa* clinical isolates, and absent in PAO1, a less pathogenic strain of *P. aeruginosa.*

The analysis of the sequencing data obtained from the pBI48 cosmid also indicated the presence of two sequences flanking the 33A9 gene which contained recognition motifs involved in cell attachment. Sequence analysis of ORF11738 (2436 bp) and ORF23228 (2565 bp), upstream and downstream of 33A9 respectively (Figure 1),

indicated the presence of RGD motifs in these two open reading frames. RGD tripeptide sequences are a characteristic eukaryotic recognition motif that binds to host cell surface integrins and have been found to be involved in bacterial adherence. By mimicking host molecules, bacterial adhesins that contain these RGD motifs can effect responses in the host that are required to promote cell-cell adhesion.

The expression of these two RGD-containing ORFs was evaluated in both 33A9 and the wild type strain PA14. Transcript levels were determined by hybridization with a radiolabeled DNA probe that corresponded to an internal region of ORF11738 and ORF23228. The data obtained for the two ORFs in the mutant 33A9 showed reduced transcript levels compared to the wild type PA14, indicating that the genes encoded by ORF11738 and ORF23228 are both regulated by 33A9. These data indicated that 33A9 plays a role as a multigene regulator responsible for the regulation of the expression of genes involved in bacterial attachment to host cell surfaces.

Table 3

ORF	Start	Stop	Length	Blast n	BlastP	Motif	Terminator	Shine-Delgarno
244c	244	35	210					
602c	602	42	561					730
214	214	792	579					
594	594	3734	3141		Conjugal transfer prtn	ATP/GTP BINDING		730
1205C	1205	987	219					
1640C	1640	1206	435					
1615C	1615	1439	177		rev transcriptase			
2929c	2929	2288	642		adhesin precursor			
3994c	3994	3818	177		outer memb. protein			
4506C	4506	3862	645			lipoprotein	4442	
4901c	4901	4668	234		atp-dep. rna helicase		4726	
10475	10475	10828	354		unk. mycobacterium			
11738	11738	14173	2436		mycobact.unk.	ATP/GTP BINDING		
14155	14155	16101	1947		DNA helicase	ATP/GTP BINDING	15915	
21421	21421	22761	1341	several P. a. genes	integrase		22982	21464
22505	22505	22657	153	t-RNAs, oprL,		prenylation		
23228	23228	26197	2970	atp dep. protease		zinc protease		
26191c	26191	23612	2580	clp proteases,	ClpB	ATP/GTP BINDING	23603	
26844c	26844	26332	513		ClpB			
26486	26486	27160	675		Memb. glycoprotein			
26857c	26857	26516	342		viral nucl.antigen			
28068c	28068	27055	1014		PilS	yabO (hypothetical)		
28118	28118	29188	1071		PilS	lipoprotein		
29382	29382	31172	1791		PilS		31186	
31247c	31247	30591	657			FABprotein		
31222	31222	32523	1302		AlgB, PilR	sigma54interaction domain	31518	
32568c	32568	32065	504			tonB (Fe receptor)	32567	
33705c	33705	32569	1137		PilR, D-AA dehydrogenase		32567; 32609	33678
34274	34274	34915	642		pilin	Nterm mrthyl (pilin)		
34916	34916	35449	534		prepilin leader			
36246	36246	36875	630	Pil genes)	pilx, pily1			
41284	41284	42234	951			sugar transport	41175; 41170	
42236c	42236	41185	1052		LYTB			

In addition, using the plant and nematode screening assays (slow- or fast-killing assays) described in Ausubel et al.

(*Methods of Screening Compounds Useful for Prevention of Infection or Pathogenicity*, USSNs 08/411,560, 08/852,927, and 08/962,750, filed on March 25, 1995, May 7, 1997, and November 3, 1997, respectively), several other mutant *Pseudomonas aeruginosa* strains were identified as having decreased virulence. The slow- and fast-killing assays utilized for these studies are described below.

Slow-killing assay. For the slow-killing assay, 10 μl of an overnight bacterial culture was spread on an NG plate (modified from NGM agar described in Sulston and Hodgkin (In: *The Nematode Caenorhabditis elegans*, W.B. Wood, ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 188, pp. 587-606): (0.35% instead of 0.25% peptone was used) and incubated at 37°C for 24 hours. After 8-24 hours at room temperature (23-25°C) each plate (3.5 cm diameter) was seeded with 40-50 hermaphrodite L4 *C. elegans* strain Bristol; for statistical purposes, 3-4 replicates per trial were carried out. Plates were incubated at 25°C, and the number of dead worms were scored every 4-6 hours. A worm was considered dead when it no longer moved when touched with an eyelash and failed to display any pharyneal pumping action. For each batch of mutants assayed, PA14 and *E. coli* OP50 were used as positive and negative controls. Any worms that died as a result of being immobilized to the wall of the plate were excluded from the analysis. In order to determine LT_{50s}, data were plotted on a graph (percentage of worms killed vs. time after exposure to test strains (hour)). A curve of the form: percentage killed = A + (1-A)/(1 + exp(B - G x log(hours after exposure))) was fitted to the data using the SYSTAT 5.2.1 computer program, where A represented the fraction of worms dying in a OP50 control experiment, and B and G are parameters which were varied to fit the curve. Once B and G have been determined, LT_{50} is calculated by the formula LT_{50} = exp(B/G) x (1 - 2 x A)^(1/G).

In developing the screen, we took advantage of two observations. First, the longer it took for the worms to be killed, the more progeny were produced. Second, early larval stages are apparently more resistant to killing by *P. aeruginosa*. This provided us with a convenient and very sensitive assay for the identification of Tn*phoA* mutants that are only slightly impaired in their pathogenic potential. These attenuated mutants would be less efficient at killing worms, and the production of progeny by survivors effectively "amplifies" even a weak defect into a readily observable phenotype. Thus, on plates

containing attenuated PA14::Tn*phoA* mutants, from the initial seeded hermaphrodites, hundreds of worms were obtained. On plates seeded with a nonpathogenic mutant, thousands of worms were seen by day five and the bacterial lawn was completely consumed, whereas none or very few live worms were found on the plates seeded with the wild-type strain. Putative nonpathogenic or attenuated mutants identified in the preliminary screen were retested, and subjected to a virulence assay to determine the *C. elegans*-killing kinetics.

Fast-killing assay. The fast-killing assay, like the slow-killing assay, is useful for identifying disease-causing microbial virulence factors. In addition, the assay is useful for identifying therapeutics that are capable of either inhibiting pathogenicity or increasing an organism's resistance capabilities to a pathogen. In preferred embodiments, the fast-killing assay is carried out using a nematode strain having increased permeability to a compound, e.g., a toxin such as colchicine. Examples of nematodes having such increased permeability include, without limitation, animals having a mutation in a P-glycoprotein, e.g., PGP-1, PGP-3, or MRP-1. Such mutant nematodes are useful in the fast-killing assay because of their increased sensitivity to toxins that is due to increased membrane permeability. This characteristic results in an assay with an increased differential between full susceptibility and full resistance to toxic compounds. The fast-killing assay may also be carried out by increasing the osmolarity of the culture medium as described below.

The fast-killing assay conditions utilized herein are as follows, 5 μl of a PA14 culture grown overnight in Kings B was spread on plates (3.5 cm diameter) containing peptone-glucose medium (PG), (1% Bacto-Peptone, 1% NaCl, 1 % glucose, 1.7% Bacto-Agar). Since the efficacy of fast-killing was found to depend on osmolarity, PG medium was modified by the addition of 0.15 M sorbitol. After spreading the bacterial culture, plates were incubated at 37°C for 24 hours and then placed at room temperature for 8-12 hours. Fifteen to twenty worms were placed on the assay plate, which was then incubated at 25°C. Each independent assay consisted of 3-4 replicates. Worm mortality was scored over time, and a worm was considered dead when it failed to respond to touch as is described above. The *E. coli* strain DH5α was used as a control for the fast-killing assays.

An analysis of these strains, together with those identified above, indicated that they fell into several different classes including the following: some mutants were less pathogenic on both plants and nematodes, whereas others were reduced in either plants or nematodes, but not both. Bacterial mutants less pathogenic in plants were defined as those which, at four days post-infiltration (DPI), had a mean maximum titer (from 5 leaf samples) of two standard deviations lower relative to wild-type within the same set of experiments. The wild-type control was necessary because the maximal level reached by wild-type at four DPI could vary as much as an order of magnitude between experiments due to the effects of minor variations in growth conditions on the plant defense responses. Similarly, a mutant was characterized as reduced in pathogenicity in worms if the mean time required to kill 50% of the worms feeding on it (LT_{50} from 3 replicates) was two standard deviations less than LT_{50} of wild-type PA14 in the same experiment.

In general, those mutant strains having reduced pathogenicity in plants included *16G12*, *25A12*, *33A9*, and *33C7*; those having reduced pathogenicity in nematodes included the *35A9, 44B1, 1G2, 8C12, and 2A8*, and those having reduced pathogenicity in plants and nematodes included *25F1*, *41A5*, *50E12*, *pho15*, *12A1*, *pho23*, *34B12*, *34H4. 3E8, 23A2, and 36A4.* Tables 4 and 5 (below) summarize the pathogenicity phenotypes of these mutant strains. Sequence analysis was carried out for each of these strains having decreased virulence due to insertional mutagenesis. The DNA sequence analyses, summarized in Tables 4 and 5, showed that both novel and known genes were identified in our screening assays. Sequences from *50E12* and *41C1* each show strong similarity to previously described open reading frames (ORFs) of unknown function in *E. coli*. Mutant *35A9* identified a *mtrR* homologue of *N. gonorrhoeae* (SwissProt P39897). Mutant 25F1 identified an operon encoding 3 proteins having identity to *orfT* of *C. tepidium*, MPK, and $DjlA_{Ec}$. Sequences from *48D9*, *35H7*, and *12A1* corresponded to the *lemA*, *gacA*, and *lasR* genes, respectively. The sequences disrupted in mutants *41A5* and *44B1* do not have significant similarity to any sequence deposited in GenBank. (The *44B1*-sequence tag is only 148 bp because and there were no sequences corresponding to the 44B1 insertion in the PAO1 database were identified). Accordingly, these sequences identify additional virulence factors. The nucleotide and

amino acid sequences obtained from these experiments are shown in Figs. 10, 11, 12, 13, 14A, 14B, 15, 16, 16A, 16B, 17, 18A, 18B, 18C, 18D, and 18E and Figs. 22, 23, 24A, 24B, 24C, 24D, 24E, 25, 26, 27, and 28.

We also carried out a battery of standard biochemical tests on Tn*phoA* mutants *41A5*, *50E12*, *41C1*, *35A9*, *48D9*, *12A1*, *44B1*, and *35H7* to assess if any contained lesions in known *P. aeruginosa* virulence factors important for mammalian pathogenicity. These tests included: a standard plate assay for sensitivity to H_2O_2, as well as standard quantitative analysis of extracellular protease, elastase, phospholipase C, and pyocyanin. Except for the following, the majority of the PA14 Tn*phoA* mutants were indistinguishable biochemically from the parent PA14 strain. Mutant *12A1* exhibited decreased elastolytic and proteolytic activities but overproduced pyocyanin. Mutant *50E12* produced 3-fold higher levels of pyocyanin than PA14. Mutant *41A5* had only about 70% of wild-type levels of proteolytic activity.

A detailed description of the DNA sequence analysis and biochemical analysis of each of these mutants identified using the slow-killing assay (described above) is now presented in the following sections.

Mutant *12A1*. The Tn *phoA* insertion in *12A1* was inserted into codon 154 of the previously described *lasR* gene of *P. aeruginosa* PA01. The phenotype of *12A1*, like other known *lasR* mutants, is pleiotropic, and includes decreased elastase and protease production. In addition *12A1* produced 2-3 times more pyocyanin than the parent PA14 strain at stationary phase. Furthermore, a *lasR* mutant expressing GFP (PA14*lasR*::GFP19-1) failed to establish itself in the worm gut as very little fluorescence was detected in *C. elegans* intestines after 48 hours of feeding.

Fig. 34A shows that the defective nematode slow-killing phenotype of *12A1* was completely restored when the *P. aeruginosa* PAO1 *lasR* gene was expressed in trans under the control of the constitutive *lacZ* promoter in strain *12A1*(pKDT17). The production of elastase was also found to be restored to wild-type levels in *12A1*(pKDT17), but not the overproduction of pyocyanin. Because the pyocyanin-overproduction phenotype was not expected, we constructed a new *lasR* mutant, *lasR*::Gm, by marker exchanging a *lasR* gene interrupted by a gentamicin cassette into the PA14 genome. The *lasR*::Gm mutant was as nonpathogenic as *12A1* (Fig. 34A),

but produced normal levels of pyocyanin, suggesting that *12A1* may harbor a second mutation that resulted in the upregulation of pyocyanin production. The result also indicated that the upregulation of pyocyanin production during the stationary phase is not related to the attenuated pathogenicity phenotype.

Mutant *pho15*. Disruption of the *dsbA* gene in *pho15* was found to be responsible for the nonpathogenic phenotypes. Fig. 24B shows the nucleotide sequence (SEQ ID NO:166) and predicted amino acid sequence (SEQ ID NO:167) of PA14 *pho15*. The pathogenicity defective phenotype of *pho15* in *C. elegans* was also found to be fully restored by constitutive expression of the *E. coli* $dsbA_{Ec}$ gene or the PA14 $dsbA_{Pa}$ gene in trans in the *pho15* background (Fig. 34B). For these experiments, the *E. coli* $dsbA_{Ec}$ gene was cloned into pUCP18 as follows. The PCR-amplified *E. coli dsbA* was cloned into the *KpnI* and *XbaI* sites of pBAD18 to form pCH3. This placed the *E. coli dsbA* under the *E. coli* arabinose promoter. A 700 bp KpnI/SphI fragment containing the *E. coli dsbA* was cloned into the *KpnI/SphI* sites of pUCP18, to make *pEcdsbA*, placing the *E. coli dsbA* under the constitutive E. coli *lacZ* promoter. pEcdsbA was subsequently used to transform PA14 and *pho15* to construct strains PA14(pEcdsbA) and *pho15*(pEcdsbA), respectively.

PA14$dsbA_{Pa}$ was constructed as follows. Based on the *dsbA* sequences of PA01 (GenBank Accession number U84726), primers TMW8 (5'-GCACTGATCGCTGCGTAGCACGGC-3'; SEQ ID NO:177) and TMW9 (5'-TGACGTAGCCGGAACGCAGGCTGC-3'; SEQ ID NO:178) were used to amplify a 1126bp fragment containing the *dsbA* gene plus 176 bp upstream of the translational start of the *dsbA* gene from genomic DNA of PA14. This fragment was cloned, using the TA cloning kit (Invitrogen), into the pCR2.1 vector to generate pCR*dsbA*. The SacI/XbaI fragment-containing dsbA was cloned into SacI/XbaI digested pUCP18 to construct pPAdsbA, placing the transcription of *dsbA* under the constitutive *lacZ* promoter. Strain *pho15*(PAdsbA) was constructed by transforming *pho15* with pPA$dsbA_{Pa}$.

Mutant *25F1*. In 25F1, Tn*phoA* was found to be inserted within codon 100 of a putative gene (*orf338*) that encodes a 338 amino acid protein, the first gene of a putative 3-gene operon. The predicted downstream genes (*orf224* and *orf252*) encode 224 and 252 amino acid proteins, respectively. GAP analysis showed that *orf338* is 28.5%

identical (37.7% similar) to *orfT* of *C. tepidum* (GenBank Accession number U58313). BLASTP of ORF224 identified mannose-1-phosphate guanylyltransferase (MPG; EC 2.7.7.13) from eukaryotes, archeabacteria, cyanobacteria, and mycobacteria, but not proteobacteria, close relatives of *P. aeruginosa.* It is not clear if ORF224 is a functional MPG since all known MPGs consist of 359-388 amino acid residues, whereas OFR224 consists of only 224 amino acid residues. ORF252 is homologous to *E. coli* $DjlA_{Ec}$. $DjlA_{Ec}$ is thought to play a role in the correct assembly, activity and/or maintenance of a number of membrane proteins, including the two-component histidine kinase signal-transduction systems.

To test if *orf338* is the gene responsible for reduced pathogenicity in worms, we compared the killing kinetics of a strain carrying *orf338* alone, *25F1*(pORF338), to wild type PA14 and 25F1 carrying vector alone. The 25Fl(pORF338) was constructed as follows.

A 1.8 kb PCR-fragment containing 482 bp upstream promoter sequence, the entire *orf338* and a truncated *orf224* was amplified (ExpandTM High Fidelity System, Boehringer Mannheim) from PA14 genomic DNA using primers F2327 (5'-CGAGGAATCCAGTCGAGGTG-3'; SEQ ID NO:179) and R4180 (5'-GCAAGATGCAGCCGAGAGTAG-3'; SEQ ID NO:180). The product was cloned into vector pCR2.1 (TA Cloning, Invitrogen) to construct plasmid pMT403C-R. The SacI/XbaI fragment from pMT403C-R, which contained the PCR product, was cloned into the SacI/XbaI of pUCP18 to construct pORF338, placing *orf338* under the control of its native promoter. *25F1* were transformed with pORF338 to make strain *25F1*(pORF338).

In addition, a strain which contained the entire operon (*orf338*, *orf224,* and $djlA_{Pa}$) was constructed as follows. A PCR strategy was used to amplify a 3.6 kb genomic fragment containing *orf338*, *orf224,* and $djlA_{Pa}$ and their upstream transcriptional sequences using primers RIF3115 (5'-GTCAGAATTCTCAGCTTGACGTTGTTGCCC-3'; SEQ ID NO:181) and RIR6757 (5'-GTCAGAATTCGACTTCTATTACCGCGACGCC-3'; SEQ ID NO:182). EcoRI sites (underlined) are present in the primers, but absent in the genomic sequence. Both strands of the PCR product were sequenced to determine the sequence of *orf338*, *orf224*,

and *djlA*$_{Pa}$ in strain PA14. The PCR EcoRI digestion product was cloned into the EcoRI site of pUCP18, and the orientation of insertion determined by restriction digest. Plasmid p3-ORFs, where *orf338*, *orf224*, and *djlA*$_{Pa}$ are under the control by its native promoter was then used to transform 25F1 to make strain *25F1*(p3-ORFs).

As is shown in Fig. 34C, strain *25F1*(pORF338) failed to complement fully the slow-killing phenotype. Strain *25F1*(p3-ORFs), which contained the entire operon (*orf338*, *orf224*, and *djlA*$_{Pa}$), also showed only partial complementation of the mutant phenotype. This result indicated that the Tn*phoA* is responsible for the pathogenicity phenotype; partial complementation may be a consequence of gene dosage. The higher mortality achieved by strain *25F1*(p3-ORFs) compared to strain *25F1*(pORF338) further suggested that the downstream genes, ORF224 and/or DjlA$_{Pa}$ may also play a role in PA14 virulence.

Fig. 24E shows the nucleotide sequence (SEQ ID NO:173) of PA14 25F1 encoding ORFT (SEQ ID NO:174), ORFU (SEQ ID NO:175), and DjlA$_{Pa}$ (SEQ ID NO:176).

Mutant *50E12*. The Tn*phoA* insertion in 50E12 was inserted within codon 39 of a predicted 759 amino acid protein that is 43% identical (54% similar) to the PtsP$_{Ec}$ protein of *E. coli*. Based on sequence analysis, *ptsP*$_{Ec}$ is predicted to encode Enzyme INtr, a 738 amino acid protein which contains an N-terminal Nif-A domain and a C-terminal Enzyme I domain; the latter functions in the phosphoenolpyruvate-dependent phosphotransferase system. It is thought the Nif-A domain serves a signal transduction function, either directly sensing small molecule signals or receiving signals from a NifL-like protein. Either mechanism may modulate the catalytic activity of the Enzyme I domain; which in turn is suggested to phosphorylate NPr (nitrogen-related HPr) and thereby regulate transcription of RpoN-dependent operons. Immediately upstream of the PA14 *ptsP*$_{Pa}$ homologue is open reading frame (*orf159*) predicted to encode a 159 amino acid protein that appears to be co-transcribed with *ptsPPa*. Fig. 24C shows the nucleotide sequence (SEQ ID NO:168) of PA14 50E12 encoding YgdP$_{Pa}$ (SEQ ID NO:169) and PtsP$_{Pa}$ (SEQ ID NO:170). ORF159 is 62.3-64.8% identical to YgdP proteins of unknown function found in *H. influenzae* (GenBank Accession number Q57045) and *E. coli* (GenBank Accession number Q46930). These proteins are closely related to invasion

protein A in *Helicobacter pylori* and *Bartonella bacilliformis*. *B. bacilliformis* invasion protein A (SwissProt Accession number P35640) is encoded by *ailA*, which when present together with an adjacent but independently transcribed gene, *ailB*, confers on *E. coli* the ability to invade human erythrocytes .

For the complementation of *50E12*, two strains were tested: *50E12*(pMT206-lac) and *50E12*(pMT206-nat). Strain *50E12*(pMT206-lac) carried plasmid pMT206-lac, where the transcription of orf159 and ptsPPa is under the control of the constitutive lacZ promoter. For strain *50E12*(pMT206-nat), the transcription of *orf159* and $ptsP_{Pa}$ is controlled only by their native promoter. Each of these strains were constructed as follows.

A 4.3 kb PCR fragment, containing the EcoRI site at both ends was amplified from genomic DNA of *P. aeruginosa* PA14 using these primers: RIF1698 (5'-GTCAGAATTCGATGTTCCAGTCCCAGATCCC-3'; SEQ ID NO:183) and RIR6002 (5'-GTCAGAATTCCAGTAGACCACCGCCGAGAG-3': SEQ ID NO:184). This fragment was cloned into the EcoRI site of pUCP18 to make pMT206-lac and pMT206-nat; their identity confirmed by restriction digest. In pMT206-lac, the transcription of *orf159* and *ptsPPa* is under the control of both the constitutive *lacZ* promoter and their native promoter. Only their native promoter controls the transcription of *orf159* and *ptsPPa* in pMT206-nat.

As is shown in Fig. 34D, both strains partially complemented the mutant phenotype, with the time required by these complemented strains to kill 100% of the worms being longer than the wild-type strain. Partial complementation was observed in the burned-mouse assay: Mortality of mice after infection by 5 x 10^5 bacteria from strain 50E12(pMT206-nat) was 39%, compared to 100% and 0% mortality when infected by the wild-type strain and 50E12, respectively. These results indicated that the putative *orf159-*$ptsP_{Pa}$ operon is involved in *P. aeruginosa* pathogenesis in nematode and mice.

Mutant *35A9*. The Tn*phoA* insertion in *35A9* is located in a putative 210 amino acid protein (encoded by *orf210*) that is most closely related (31.5% identity) to the *N. gonorrhoeae* MtrRNg protein, which belongs to the TetR family of helix-turn-helix containing bacterial transcription regulation proteins. ORF210 is adjacent to, and divergently transcribed from, three genes that are homologous to components of the

energy dependent efflux (EDE) system in *P. aeruginosa*. Analyses of sequences from PA01 showed that together, these four genes defined a novel energy dependent efflux (EDE) system in *P. aeruginosa*. The other EDE systems in *P. aeruginosa* described previously are the *mexR*, *mexA-mexB-oprK* system, the *nfxB*, *mexC-mexD-oprJ* system and the *nfxC*, *mexE-mexF-oprN* system. Fig. 24D shows the nucleotide sequence (SEQ ID NO:171) of PA14 35A9 encoding $mtrR_{Pa}$ (SEQ ID NO:172).

Mutants *37H7* and *1D7*. Analysis of the IPCR product from mutant *37H7* showed that there is a Tn*phoA* insertion within codon 188 of the 214 amino acid GacA protein. DNA blot analysis showed that *1D7* also contained an insertion in the *gacA* gene.

Mutant *48D9*. Tn*phoA* is inserted between codon 491 and 492 of the 925 amino acid LemA-homologue, a sensor kinase belonging to a family of bacterial two-component regulators . The cognate response regulator of LemA in *P. syringae* is GacA and GacA + LemA have been shown to affect the expression of a variety number of virulence factors .

Mutant *41C1*. Tn*phoA* is inserted in the AefA-homologue of the putative *E. coli* integral membrane protein (SwissProt P77338) in mutant *41C1*. It is a member of the 30-40 kD UPF0003 protein family (PROSITE PDOC00959). In addition to *E. coli*, it is also present *Synechocystis* strain PCC 6803 and *Methanococcus jannaschii*.

In addition, strains *pho34B12*, *3E8*, *8C12*, *1G2*, *35A9, and 23A2*, were also found to have a phenazine-minus mutant phenotype. Moreover, *pho34B12*, *3E8*, *8C12*, and *1G2* mutants were found to be reduced in pigment production. An additional mutant, *6A6*, was also identified having reduced pigment. The characteristic color of *P. aeruginosa* strains has been attributed to a group of tricyclic secondary metabolites collectively known as phenazines, the most extensively characterized of which is the blue-green pigment, pyocyanin (1-hydroxy-5-methyl phenazine). In order to test whether the reduction of pigmentation in the bacterial mutants was at least in part due to the reduction in pyocyanin, levels of this pigment were quantified in wild type PA14 as well as in all the mutants obtained using the fast-killing assay. The results of this analysis showed that the *pho34B12*, *3E8*, *8C12*, *1G2, and 6A6* mutants that had a reduced pigment

phenotype were also reduced in pyocyanin production, with levels ranging from 10 to 50% of the wild type strain. The other mutants, *13C9*, *23A2*, and *36A4* had levels of pyocyanin comparable with the wild type strain.

In addition, the sequence interrupted by the Tn*phoA* mutation in *3E8* was found to predict a protein with homology to the *phzB* gene from *Pseudomonas fluorescens*, that is part of an operon involved in the production of the secondary metabolite, phenazine (GenBank Accession number: L48616). The *phzB* gene also has a homolgue in *Psuedomonas aureofaciens*, referred to as *phzY*. (GenBank Accession number AF007801). Using the sequence tag, a cosmid (1G2503), containing this region in the *Pseudomonas aeruginosa* database was identified, that contains both the *phzA* and *phzB* genes, as well as other genes that are thought to play a role in phenazine biosynthesis, the *pcnC* and *D* genes (GenBank Accession number AF005404). Four of these strains, *34B12*, *3E8*, 23A12, and 35A9, were examined for pathogenicity in the mouse-burn assay. Surprisingly, these experiments showed that the phenazine defective strains have reduced pathogenesis, indicating that the genes interrupted by the Tn*phoA* insertions are mammalian virulence factors. The nucleotide and deduced amino acid sequences, including sequence tags, for these strains are shown in Figs.7-9, 13, 14A, 14B, 15, 16A, 16B, 17, 18A, 18B, 18C, 18D, 18E, 22, 24A, 24B, 24C, 24D, 24E, and 33. In addition, Figs. 25 and 26 show the nucleotide sequence of the *phnA* and *phnB* genes of *Pseudomonas aeruginosa* and the deduced amino acid sequence of PHNA, respectively.

A detailed description of the DNA sequence and biochemical analyses of each of the mutants identified using the fast-killing assay (described above) is now presented in the following sections.

Mutants *36A4*, *23A2*, and *13C9*. The DNA sequence tags obtained from all three of the mutants that produced wild type levels of pyocyanin, had homologies to known genes in Pseudomonads. Mutant *36A4* contained Tn*phoA* inserted into a gene homologous to *hrpM*, previously identified as a locus controlling pathogenicity in the plant pathogen *Pseudomonas syringae* (Mills and Mukhopadhyay, In: *Pseudomonas*: biotransformations, pathogenesis, and evolving technology, S. Silver, A.M. Chakrabarty, B. Iglewiski, and S. Kaplan, eds., American Society for Microbiology, 1990, pp. 47-57, Mukhopadhyay et al., *J. Bacteriol.* 170:5479-5488, 1988); GenBank Accession number

140793). This locus also has homology to the *E. coli mdoH* gene, which encodes an enzyme involved in the biosynthesis of periplasmic glucans (Loubens et al., *Mol. Microbiol.* 10:329-340, 1993; GenBank Accession number X64197). The Tn*phoA* insertion in mutant *23A2* was inserted into a gene previously identified in *P. aeruginosa* strain PAO1 as *mexA* (Poole et al., *Mol. Microbiol.* 10:529-544, 1993; GenBank Accession number L11616). The product of *mexA*, predicted to be a cytoplasmic-membrane-associated lipoprotein, likely functions together with the products of the other two genes contained in the same operon, *mexB* and *oprM*, as a non-ATPase efflux pump with broad substrate specificity (Li et al., *Antimicrob. Agents. Chemother.* 39:1948-1953, 1995). Sequence analysis of the DNA flanking the third mutant that was wild type for pigment production, *13C9*, showed that it corresponded to another previously known gene in *P. aeruginosa* strain PAO1, *orp* (GenBank Accession number U54794). *Orp,* or osmoprotectant-dependent regulator of phospholipase C, was identified as a factor controlling the expression of the pathogenicity factor *PlcH*, one of the two isoforms of phosholipase C produced by *P. aeruginosa* (Sage et al.,Mol. Microbiol. 23: 43-56, 1997).

Mutants *1G2* and *8C12*. Molecular analysis of two of the non-pigmented mutants *1G2* and *8C12* showed that they contained insertions into novel genes, although DNA flanking the *1G2* insertion contained a motif characteristic of histidine sensor kinases. This gene was not present in the PAO1 genome database. Although the *8C12* sequence tag identified a homologous gene in the PAO1 database, no significant motifs were found within this gene.

Mutants *3E8* and *6A6*. Two mutants, *3E8* and *6A6*, contained Tn*phoA* insertions into the same gene, which was homologous to the previously identified *phzB* gene in *P. fluorescens* strain 2-79 (GenBank Accession number AF007801) and *phzY* in *P. aureofaciens*, strain 30-84 (GenBank Accession number L48616). These two mutants contained the Tn*phoA* insertion in exactly the same position, however, they were independent isolates since they were obtained from two different mutant libraries. Although *phzB* and *phzY* contained no identifiable sequence motifs, they were present in operons known to regulate production of phenazine-1-carboxylate (PCA) in both *P. fluorescens* and *P. aureofaciens* (Mavrodi et al., *J. Bacteriol.* 180:2541-2548, 1998).

Mutant pho34A12. DNA flanking the Tn*phoA* insertion in the final non-pigmented mutant *pho34B12*, was previously cloned and shown to be a novel locus as described *infra*. Interestingly, this insertion is immediately downstream of the phenazine biosynthetic genes, *phnA* and *phnB*, as identified in *P. aeruginosa* strain PAO1 (Essar et al., *J. Bacteriol.* 172:884-900, 1990).

Phenazines are required for fast killing of *C. elegans*

The isolation of both pigmented and non-pigmented mutants in the fast-killing screen indicated that the fast-killing process involved more than one factor. However, the molecular analysis of the *3E8* and *6A6* mutants (containing insertions in an operon known to regulate phenazine production) strongly suggested that phenazines represented one class of toxin that mediate fast killing. In order to directly test this hypothesis, an additional mutation, ΔphnA phnB, was generated and studied as follows.

The phenazine biosynthetic genes *phnA* and *phnB* (Essar et al., *J. Bacteriol.* 172:884-900, 1990) genes lie upstream of the previously characterized *pho34B12* Tn*phoA* insertion in PA14; GenBank Accession number AF031571). A 3.7 kb EcoRI fragment corresponding to the wild type sequence of this region (from the plasmid pLGR34) was subcloned into pBluescript SK/+ to yield Bs34B12. This plasmid contained 944 bp of *phnA* (full length of 1591 bp), the entire *phnB* (600 bp) gene and 1.7 kb of downstream sequences. The missing 605 bp of *phnA* and 405 bp upstream were amplified using PCR from genomic PA14 DNA with the oligonucleotide primers PHNA3 (5'-GGTCTAGACGAACTGAGCGAGGAG-3'; SEQ ID NO:185) and PHNA2 (5'-GCCTGCAGGCGTTCTACCTG-3'; SEQ ID NO:186). The primers were based on the sequence of the previously cloned *phnA* and *phnB* genes from *P. aeruginosa* strain PAO1 (Essar et al., *J. Bacteriol.* 172:884-900, 1990, GenBank Accession number M33811). The 1010 bp amplified sequence was subcloned into the PstI sites of *pBs34B12* to give the construct, pBs34B12phnA. An in-frame deletion within *phnA*, *phnB* was generated by replacing 2.6 kb of the wild type sequence of the genes with a 1 kb fragment (Fig. 35) amplified by PCR using the primers PHNDEL1 (5'-GGCTGCAGTGATTGACTGAGCGTCTGCTGGAGAACG-3'; SEQ ID NO:187) and PHNDEL2 (5'-GGGAAGCTTCGTCTAGAATCACGTGAACATGTTCC-3': SEQ ID NO:188) to yield the plasmid pBs34b12phndel. A 1.8 kb XbaI fragment containing

the *phnAphnB* in-frame deletion was subcloned into the positive-sucrose-selection suicide vector pCVD442 (Donnenberg and Kaper, *Infect. Immun.* 59:4310-4317, 1991). The resulting construct, pCVD34B12phndel, was used to introduce the disrupted *phnA*, *phnB* genes into the wild-type PA14 genome by homologous recombination resulting in the mutant PA14 Δ*phnAphnB*. DNA restriction and DNA blot analyses using DNA from the parental PA14 and derivative PA14 Δ*phnAphnB* strains were undertaken in order to verify that the mutant contained the desired deletion.

Although little is known about the nature of the enzymes that catalyze the formation of phenazines in *P. aeruginosa* and related Pseudomonads, the conversion of chorismate to anthranilate is thought to be a key step in the pathway (Fig. 35A). In *P. aeruginosa* strain PAO1, this step is most-likely catalyzed by the anthranilate synthase encoded by the *phnA* and *phnB* genes, since mutations in these genes result in decreased production of the phenazine pyocyanin (Essar et al., *J. Bacteriol.* 172:884-900, 1990). The *phnA* and *phnB* genes were cloned from PA14 and a ΔphnAphnB mutant containing a 1602 bp deletion in these genes was generated (Fig. 35B). Importantly, this mutation was designed to be non-polar and therefore did not affect the two ORFs shown to be directly downstream of *phnA* and *phnB* (infra). Measurement of pyocyanin in the ΔphnAphnB mutant showed that it generated only 10% of wild type levels, confirming that *phnA* and *phnB* are involved in pyocyanin production in strain PA14 just as in PAO1. Assays conducted using ΔphnAphnB revealed that this strain was severely reduced in fast killing. As seen in Figure 35C, less than 5% of the worms were dead three hours after exposure to ΔphnAphnB in contrast to almost 100% that were exposed to the wild type strain. The ΔphnAphnB strain behaved in a manner similar to the other phenazine mutant, *3E8*, which served as the control for an attenuated mutant in this experiment. These results demonstrated that phenazines are required for the fast killing of *C. elegans*.

To discover whether the bacterial factors that mediated fast killing are relevant to pathogenesis in other hosts, the fast-killing mutants were tested for virulence in the *Arabidopsis* leaf infiltration model as well as the mouse full thickness skin burn model (*infra*). Five fast-killing mutants were tested for growth over the course of four days in *Arabidopsis* leaves as a quantitative measure of their pathogenicity and also in the mouse full thickness skin burn model. As shown in Tables 4 and 5, the maximal level of growth

in *Arabidopsis* leaves on the fourth day postinfection was significantly lower for 2 of the phenazine mutants, *3E8* and *8C12*. In the mouse model these two mutants caused significantly less mortality than the wild type strain with a $P < 0.05$ when an inoculum of 5×10^5 cells was used. The third phenazine mutant *1G2*, was not significantly different from the wild type strain in either the plant or the mouse models.

Both the *hrpM* mutant, *36A4*, and the *mexA* mutant, *23A2*, were severely debilitated in growth in *Arabidopsis* leaves, indicating a strong pathogenicity defect in this model. In the mouse model, mutant *36A4*, had a dramatic effect causing no mortality at the dose tested. In contrast, the *mexA* mutant, *23A2* was only marginally affected. These results demonstrated that the fast killing screen is extremely effective at identifying genes required for pathogenesis in both plants and mice, and further, provide the first *in vivo* demonstration that phenazines are required for pathogenesis in these two hosts.

We also note that we have identified a regulator, *phzR*, of the *phz* operon. Fig. 18E shows the nucleotide sequence (SEQ ID NO:164) and predicted partial amino acid sequence (SEQ ID NO:165) of PA14 *phzR*.

Phenazines and pathogenesis

PA14 mutants reduced in fast killing also affected pigment synthesis. Our molecular analysis revealed that the association between pigment production and pathogenesis was not simply due to the coordinate regulation of pigmentation and toxin production by regulatory factors. Instead we found that mutations in phenazine biosynthetic genes were reduced in virulence, strongly implicating phenazines as toxins in the fast-killing process. Phenazines, tri-cyclic pigmented compounds that give Pseudomonads their characteristic colors (Turner and Messenger, *Adv. Microb. Physiol.* 27:211-273, 1986), are secondary metabolites thought to increase the survival of organisms under competitive conditions (Maplestone et al., *Gene* 115:151-157, 1992). Although the repertoire of phenazines produced by PA14 is unknown, *P. aeruginosa* strain PAO1 produces at least six different phenazines, including the well characterized blue-green pigment pyocyanin. Phenazines including pyocyanin, have been demonstrated to have antibiotic action against several species of bacteria, fungi, and protozoa, a quality attributed to their redox active. In their highly-reactive reduced state, phenazines have been described to undergo redox cycling in the the presence of various reducing agents or

molecular oxygen resulting in the formation of superoxide and hydrogen peroxide (Hassan and Fridovich, *J. Bacteriol.* 141:1556-163, 1980). *In vitro*, these moderately cytotoxic oxygen radicals can be converted by an iron catalyst to the highly cytotoxic hydroxyl radical (Britigan et al., *J. Clin. Invest.* 90:2187-2196, 1992). Formation of reactive oxygen species by phenazines is also thought to contribute to their cytotoxic effects observed on eukaryotic cells in vitro. These effects include the inhibition of mammalian cell respiration, the disruption of ciliary beating, and immunomodulatory effects such as stimulation of the inflammatory response, inhibition of lymphocyte proliferation and alteration of the T lymphocyte response to antigens.

The biosynthetic pathways leading to the production of phenazines in *P. aeruginosa* have been poorly defined making it difficult to identify the steps in the pathway blocked by the PA14 mutants defective in phenazine production. However, the transposon insertion in two mutants, *3E8* and *6A6*, disrupted a gene with homology to *phzB*, which was previously characterized as being involved in phenazine production in the related Pseudomonads, *P. fluorescens,* and *P. aureofaciens*. In *P. fluorescens*, *phzB* was shown to be part of a seven gene operon (*phzA-G*) involved in the production of phenazine-1-carboxylic acid. Comparison of this operon in *P. flourescens* and *P. aureofaciens* showed that the two were highly homologous, suggesting that pathways leading to phenazine production are conserved in fluorescent Pseudomonads (Mavrodi et al., *J. Bacteriol.* 180:2541-2548, 1998). Although the DNA flanking the *phzA* and *phzB* genes has only been partially sequenced in *P. aeruginosa* strain PA14, our analysis suggests that the region shares a conserved structure with the *P. fluorescens phzA-F* operon. The predicted translated products of the *phzA* and *phzB* genes from PA14 and *P. fluorescens* share 68 and 74% identity, respectively. In addition, a region containing *phzA-F*-like genes is present in *P. aeruginosa* strain PAO1, and the predicted translated products of these genes exhibited between 69 to 85% identity with their *P. fluorescens* homologs (GenBank Accession number AF005404). Extrapolating from the role of the *phz* operon in *P. fluorescens* and *P. aureofaciens*, the isolation of PA14 *phzB* mutants that are defective in fast killing strongly suggested that phenazines are involved in this process. The hypothesis that phenazines, including pyocyanin, are one of the mediators of fast killing was further tested by the non-polar disruption of the genes, *phnA* and *phnB*,

which encode the two subunits of an anthranilate synthase, previously shown to be specifically involved in phenazine synthesis in *P. aeruginosa* strain PA01 (Essar et al., *J. Bacteriol.* 172:884-990, 1990). Consistent with a role in phenazine biosynthesis, deletion of the *phnA* and *phnB* genes in PA14 severely reduced pyocyanin production. Furthermore, the ΔphnAphnB mutant was defective in fast killing, demonstrating the critical role of phenazines in this process.

The role of phenazines in pathogenesis was also examined in *Arabidopsis* and mice. The two independent mutants containing insertions within the *phzB* gene, *3E8,* and *6A6*, were dramatically reduced in pathogenicity in both the *Arabidopsis* leaf infiltration model as well as the mouse full thickness skin burn model (Tables 4 and 5), suggesting that phenazines are multi-host pathogencity factors. It is interesting to note that many of the other mult-host pathogenicity factors identified in this and our previous studies are likely to be involved in the production of several other virulence factors and are not effectors, or molecules that directly interact with the host (described infra). Thus, phenazines represent the only known class of multi-host pathogenicity effectors that we have identified. These findings are also significant since despite intensive *in vitro* analyses of phenazines, the physiological significance of their production and their role in *P. aeruginosa* infections remains controversial, and prior to this study there has been no demonstration of their role *in vivo*.

Fast killing is multifactorial

Analysis of fast-killing mutants that generated wild-type levels of pigments showed that although phenazines were essential mediators of fast killing, other factors were involved in this process. Molecular analysis of one such mutant, *23A2*, revealed that the transposon was inserted into a gene previously identified in *P. aeruginosa* strain PAO1 as *MexA*, which is part of the 3 gene operon *MexA, B, OprM* (Poole et al., *Mol. Microbiol.* 10:529-544, 1993). The products of these genes are localized to the cytoplasmic (*MexA, MexB*) and outer membranes (*OprM*) where they are proposed to function as a non-ATPase broad-specificity efflux pump (Li et al., *Antimicrob. Agents Chemother.* 39:1948-1953,1995). Originally identified due to its contribution to the process of multi-drug resistance in *P. aeruginosa*, this pump is thought to play a general role in the export of secondary metabolites, although its natural substrates remain

unknown (Poole, *Antimicrob. Agents Chemother.* 34:453-456, 1994). The defect of *mexA* mutant in fast killing, a process mediated by diffusible toxins, is most-likely due to the lack of export of one or more factors involved in this process. Since the *mexA* mutant was pigmented, phenazines are not likely to be a substrate for the pump. In addition to its defect in fast killing, the *mexA* mutant was marginally reduced in pathogenicity in the mouse model and severely debilitated in the *Arabidopsis* leaf infiltration model. Although the lack of export of specific virulence factors could explain these defects, an additional model is that the *mexA* mutant bacteria are unable to protect themselves against host defense factors generated in response to the bacterial infection. Such a protective function has been demonstrated for the *sap* genes, which encode proteins related to ATP binding cassette (ABC) transporters and mediate resistance to host antimicrobial peptides in the human pathogen, *Salmonella typhimurium*, as well as in the phytopathogen, *Erwinia chrysanthemi* (Taylor, *Plant Cell* 10:873-875, 1998).

A second mutant identified in the screen, *36A4*, contained a transposon insertion into a gene with homology to *E. coli MdoH*, which is part of the *mdoGH* operon. In *E. coli*, the products of this operon are involved in the synthesis of membrane-derived oligosaccharides (MDO) or linear, periplasmic glucans (Loubens et al., *Mol. Microbiol.* 10:329-340, 1993). A similar locus, termed *hrpM* is present in the plant pathogen *Pseudomonas syringae* pv. *syringae* (Mukhopadhyay et al., *J. Bacteriol.*170:5479-5488, 1988), originally identified since mutations within this locus abolish both the development of disease symptoms on host plants as well as the hypersensitive response in non-host plants (Anderson and Mills, *Phytopath.*75:104-108, 1985). Periplasmic glucans have also been found in a wide range of gram-negative bacteria, where diverse, albeit poorly understood functions have been assigned to them. In addition to being essential virulence factors in *P. syringae*, other functions include the adaptation to hypoosmotic environments, and cell signaling leading to the recognition of eukaryotic hosts by species of Rhizobium and Agrobacterium (Kennedy, In: *Escherichia and Salmonella*, F.C. Neidardt, ed, American Society for Microbiology Press, Washington, D.C., pp. 1064-1071, 1996). However, despite being present in the periplasm of several animal pathogens such as *Salmonella* and *Klebsiella*, until this study, which shows that *P. aeruginosa* carrying a mutation in an *mdoH*-like locus is severely reduced in

pathogenicity in a mouse model, periplasmic glucans have not been shown to play a role in the infection of animal hosts.

Table 4

Summary for Pathogenicity of *P. aeruginosa* strain UCBPP-PA14 mutants on various hosts

	Pathogenicity Phenotypes					
Strain Isolation Number	Strain Name	Growth in Arabidopsis Leaf[b]	Ability to kill C. elegans[c]	% Mouse Mortality $5x10^{5}$[d]	Gene Identity
PA14	PA14	5.5×10^{7}	+	100	
rep (reduced pathogenicity in plants)					
16G12	*rep1*	2.3×10^{5}	+	100	no matches
49H2	*rep2*	1.2×10^{6}	+	63	not sequenced
25A12	*rep3*	1.7×10^{6}	+	75	no matches
33A9	*rep4*	5.1×10^{6}	+	0	no matches
33C7	*rep5*	8.4×10^{5}	+	0	no matches
ren (reduced pathogenicity in nematodes)					
35A9[g]	*ren1*	5.7×10^{7}	-	55	*mtrR*
44B1	*ren2*	5.4×10^{7}	-	56	no matches
1G2[f,g,h]		NT	-	NT	no matches
8C12[f,g,h]		NT	-	NT	no matches
2A8[f,h]		NT	-	NT	no matches
rpn (reduced pathogenicity in plants and nematodes)					
25F1	*rpn1*	1.5×10^{4}	-	20	*orfT*
35H7[e]	*rpn2*	1.2×10^{4}	-	NT[e]	*gacA*
41A5	*rpn3*	1.3×10^{4}	-	100	no matches
41C1	*rpn4*	2.4×10^{5}	-	85	*aefA*
50E12	*rpn5*	2.0×10^{5}	-	0	*ptsP*
pho15	*rpn6*	3.9×10^{4}	-	62	*dsbA*
12A1	*rpn7*	1.7×10^{6}	-	50	*lasR*
pho23	*rpn8*	6.4 x 104	-	5	no matches
34B12[g,h]	*rpn11*	4.0×10^{4}	-	50	dst* of phnB
34H4	*rpn12*	3.8×10^{6}	-	50	no matches

25F1	*rpn1*	1.5×10^4	-	20	*orfT*
3E8[g,h]	*rpn 13*	1×10^6	-	12.5	phzB
23A2[h]	*rpn14*	1.7×10^5	-	71	mexA
36A4[h]	*rpn15*	4×10^4	-	0	hrpN

[b]CFU/cm^2 leaf area of bacterial counts at four days after inoculation of 10^3 bacteria: means of four to five samples. Mutants are defined as less pathogenic when the means of four to five samples. Mutants are defined as less pathogenic when the mean CFU/cm^2 leaf area of bacterial counts is 2 standard deviation lower relative to wild-type within the same set of experiments.

[c]A mutant is considered attenuated in nematode pathogenicity (-) if the mean time required to kill 50% of the worms feeding on it (LT_{50} from 3 replicates) is two standard deviations less than the LT_{50} of parental UCBPP-PA14 in the same experiment; for calculations of LT_{50} see Materials and Methods.

[d]Six-week old male ARK/J inbred strain mice (from Jackson Laboratories), weighing between 20 to 30 gm were injected with 5×10^5 cells as described by Stevens et al., *J. of Burn Care and Rehabil.*15:232-235, 1994. The number of animals that died of sepsis was monitored each day for ten days.

[e]Two other independently isolated *gacA* mutants are ID7(*rpn9*) and 33D11(*rpn10*). Mutant rpn9 has been tested on mice and showed 50% mortality

[f] tested only in nematodes

[g] phenazine-defective mutants

[h] mutants defective in fast killing, not affected in slow killing

dst*=downstream

Table 5

Pathogenicity of PA14 Fast Killing Mutants in Plants and Mice

Strain	Growth in *Arabidopsis* leaves[a]	% Mouse Mortality (n) 5×10^5[b]	Gene Identity
PA14	7×10^8	100 (>16)	
1G2	3×10^7	100 (8)	no matches, contains histidine kinase motif
3E8[c], *6A6*	3×10^5	18 (16)	*phzB*
8C12	5×10^5	63 (8)	no matches
23A2	1.2×10^4	85 (16)	*mexA*
36A4	2×10^4	0 (16)	*hrpM*

[a] CFU/cm^2 leaf area of bacterial counts at five days post-inoculation with 10^3 bacteria. Values represent means of four to five samples. Mutants are defined as less pathogenic when the mean value of bacterial counts is two standard deviations lower than the wild type within the same experimental set.

[b] Six-week old male AKR/J inbred mice (from Jackson laboratories), weighing between 20 to 30 gm were injected with 5×10^5 bacterial cells. (n) is the total number of mice injected. The number of mice that died of sepsis was monitored daily for seven days.

[c] *3E8* and *6A6* are independently generated mutants that contain Tn*phoA* inserted in exactly the same location. The numbers reported are those obtained using *3E8*. Similar results were obtained with *6A6* (data not shown).

Isolation of Additional Virulence Genes

Based on the nucleotide and amino acid sequences described herein, the isolation of additional coding sequences of virulence factors is made possible using standard strategies and techniques that are well known in the art. Any pathogenic cell can serve as the nucleic acid source for the molecular cloning of such a virulence gene, and these sequences are identified as ones encoding a protein exhibiting pathogenicity-associated structures, properties, or activities.

In one particular example of such an isolation technique, any one of the nucleotide sequences described herein may be used, together with conventional screening methods of nucleic acid hybridization screening. Such hybridization techniques and screening procedures are well known to those skilled in the art and are described, for example, in Benton and Davis (*Science* 196:180, 1977); Grunstein and Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 1997); Berger and Kimmel (*supra*); and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York. In one particular example, all or part of the 33A9 sequence (described herein) may be used as a probe to screen a recombinant plant DNA library for genes having sequence identity to the 33A9 gene (Figs. 5 and 6A-B). Hybridizing sequences are detected by plaque or colony hybridization according to standard methods.

Alternatively, using all or a portion of the amino acid sequence of the 33A9 polypeptide, one may readily design 33A9-specific oligonucleotide probes, including degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either DNA strand and any appropriate portion of the 33A9 sequence (Figs. 5 and 6A-B; SEQ ID NOs:102 and 103, respectively) of the 33A9 protein. General methods for designing and preparing such probes are provided, for example, in Ausubel et al. (supra), and Berger and Kimmel, *Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York. These oligonucleotides are useful for 33A9 gene isolation, either through their use as probes capable of hybridizing to 33A9 complementary sequences or as primers for various amplification techniques, for example, polymerase chain reaction

(PCR) cloning strategies. If desired, a combination of different, detectably-labelled oligonucleotide probes may be used for the screening of a recombinant DNA library. Such libraries are prepared according to methods well known in the art, for example, as described in Ausubel et al. (*supra*), or they may be obtained from commercial sources.

As discussed above, sequence-specific oligonucleotides may also be used as primers in amplification cloning strategies, for example, using PCR. PCR methods are well known in the art and are described, for example, in *PCR Technology*, Erlich, ed., Stockton Press, London, 1989; *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc., New York, 1990; and Ausubel et al. (*supra*). Primers are optionally designed to allow cloning of the amplified product into a suitable vector, for example, by including appropriate restriction sites at the 5' and 3' ends of the amplified fragment (as described herein). If desired, nucleotide sequences may be isolated using the PCR "RACE" technique, or Rapid Amplification of cDNA Ends (see, e.g., Innis et al. (*supra*)). By this method, oligonucleotide primers based on a desired sequence are oriented in the 3' and 5' directions and are used to generate overlapping PCR fragments. These overlapping 3'- and 5'-end RACE products are combined to produce an intact full-length cDNA. This method is described in Innis et al. (*supra*); and Frohman et al., *Proc. Natl. Acad. Sci. USA* 85:8998, 1988.

Partial virulence sequences, e.g., sequence tags, are also useful as hybridization probes for identifying full-length sequences, as well as for screening databases for identifying previously unidentified related virulence genes. For example, the sequences of 36A4, 25A12, and 33C7 were expanded to those encompassed by contigs 2507, 1126, and 1344, respectively.

Confirmation of a sequence's relatedness to a pathogenicity polypeptide may be accomplished by a variety of conventional methods including, but not limited to, functional complementation assays and sequence comparison of the gene and its expressed product. In addition, the activity of the gene product may be evaluated according to any of the techniques described herein, for example, the functional or immunological properties of its encoded product.

Once an appropriate sequence is identified, it is cloned according to standard

methods and may be used, for example, for screening compounds that reduce the virulence of a pathogen.

Polypeptide Expression

In general, polypeptides of the invention may be produced by transformation of a suitable host cell with all or part of a polypeptide-encoding nucleic acid molecule or fragment thereof in a suitable expression vehicle.

Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems may be used to provide the recombinant protein. The precise host cell used is not critical to the invention. A polypeptide of the invention may be produced in a prokaryotic host (e.g., *E. coli*) or in a eukaryotic host (e.g., *Saccharomyces cerevisiae*, insect cells, e.g., Sf21 cells, or mammalian cells, e.g., NIH 3T3, HeLa, or preferably COS cells). Such cells are available from a wide range of sources (e.g., the American Type Culture Collection, Rockland, MD; also, see, e.g., Ausubel et al., *supra*). The method of transformation or transfection and the choice of expression vehicle will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel et al. (*supra*); expression vehicles may be chosen from those provided, e.g., in *Cloning Vectors: A Laboratory Manual* (P.H. Pouwels et al., 1985, Supp. 1987).

One particular bacterial expression system for polypeptide production is the *E. coli* pET expression system (Novagen, Inc., Madison, WI). According to this expression system, DNA encoding a polypeptide is inserted into a pET vector in an orientation designed to allow expression. Since the gene encoding such a polypeptide is under the control of the T7 regulatory signals, expression of the polypeptide is achieved by inducing the expression of T7 RNA polymerase in the host cell. This is typically achieved using host strains which express T7 RNA polymerase in response to IPTG induction. Once produced, recombinant polypeptide is then isolated according to standard methods known in the art, for example, those described herein.

Another bacterial expression system for polypeptide production is the pGEX expression system (Pharmacia). This system employs a GST gene fusion system which is

designed for high-level expression of genes or gene fragments as fusion proteins with rapid purification and recovery of functional gene products. The protein of interest is fused to the carboxyl terminus of the glutathione S-transferase protein from *Schistosoma japonicum* and is readily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Fusion proteins can be recovered under mild conditions by elution with glutathione. Cleavage of the glutathione S-transferase domain from the fusion protein is facilitated by the presence of recognition sites for site-specific proteases upstream of this domain. For example, proteins expressed in pGEX-2T plasmids may be cleaved with thrombin; those expressed in pGEX-3X may be cleaved with factor Xa.

Once the recombinant polypeptide of the invention is expressed, it is isolated, e.g., using affinity chromatography. In one example, an antibody (e.g., produced as described herein) raised against a polypetide of the invention may be attached to a column and used to isolate the recombinant polypeptide. Lysis and fractionation of polypeptide-harboring cells prior to affinity chromatography may be performed by standard methods (see, e.g., Ausubel et al., *supra*).

Once isolated, the recombinant protein can, if desired, be further purified, e.g., by high performance liquid chromatography (see, e.g., Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, eds., Work and Burdon, Elsevier, 1980).

Polypeptides of the invention, particularly short peptide fragments, can also be produced by chemical synthesis (e.g., by the methods described in *Solid Phase Peptide Synthesis*, 2nd ed., 1984 The Pierce Chemical Co., Rockford, IL).

These general techniques of polypeptide expression and purification can also be used to produce and isolate useful peptide fragments or analogs (described herein).

Antibodies

To generate antibodies, a coding sequence for a polypeptide of the invention may be expressed as a C-terminal fusion with glutathione S-transferase (GST) (Smith et al., *Gene* 67:31-40, 1988). The fusion protein is purified on glutathione-Sepharose beads, eluted with glutathione, cleaved with thrombin (at the engineered cleavage site), and purified to the degree necessary for immunization of rabbits. Primary immunizations are

carried out with Freund's complete adjuvant and subsequent immunizations with Freund's incomplete adjuvant. Antibody titres are monitored by Western blot and immunoprecipitation analyses using the thrombin-cleaved protein fragment of the GST fusion protein. Immune sera are affinity purified using CNBr-Sepharose-coupled protein. Antiserum specificity is determined using a panel of unrelated GST proteins.

As an alternate or adjunct immunogen to GST fusion proteins, peptides corresponding to relatively unique immunogenic regions of a polypeptide of the invention may be generated and coupled to keyhole limpet hemocyanin (KLH) through an introduced C-terminal lysine. Antiserum to each of these peptides is similarly affinity purified on peptides conjugated to BSA, and specificity tested in ELISA and Western blots using peptide conjugates, and by Western blot and immunoprecipitation using the polypeptide expressed as a GST fusion protein.

Alternatively, monoclonal antibodies which specifically bind any one of the polypeptides of the invention are prepared according to standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256:495, 1975; Kohler et al., *Eur. J. Immunol.* 6:511, 1976; Kohler et al., *Eur. J. Immunol.* 6:292, 1976; Hammerling et al., In *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, NY, 1981; Ausubel et al., supra). Once produced, monoclonal antibodies are also tested for specific recognition by Western blot or immunoprecipitation analysis (by the methods described in Ausubel et al., supra). Antibodies which specifically recognize the polypeptide of the invention are considered to be useful in the invention; such antibodies may be used, e.g., in an immunoassay. Alternatively monoclonal antibodies may be prepared using the polypeptide of the invention described above and a phage display library (Vaughan et al., *Nature Biotech* 14:309-314, 1996).

Preferably, antibodies of the invention are produced using fragments of the polypeptide of the invention which lie outside generally conserved regions and appear likely to be antigenic, by criteria such as high frequency of charged residues. In one specific example, such fragments are generated by standard techniques of PCR and cloned into the pGEX expression vector (Ausubel et al., *supra*). Fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described

in Ausubel et al. (*supra*). To attempt to minimize the potential problems of low affinity or specificity of antisera, two or three such fusions are generated for each protein, and each fusion is injected into at least two rabbits. Antisera are raised by injections in a series, preferably including at least three booster injections.

Antibodies against any of the polypeptides described herein may be employed to treat bacterial infections.

Screening Assays

As discussed above, we have identified a number of *P. aeruginosa* virulence factors that are involved in pathogenicity and that may therefore be used to screen for compounds that reduce the virulence of that organism, as well as other microbial pathogens. For example, the invention provides methods of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of a polypeptide or the gene expression of a nucleic acid sequence of the invention. The method of screening may involve high-throughput techniques.

Any number of methods are available for carrying out such screening assays. According to one approach, candidate compounds are added at varying concentrations to the culture medium of pathogenic cells expressing one of the nucleic acid sequences of the invention. Gene expression is then measured, for example, by standard Northern blot analysis (Ausubel et al., *supra*), using any appropriate fragment prepared from the nucleic acid molecule as a hybridization probe. The level of gene expression in the presence of the candidate compound is compared to the level measured in a control culture medium lacking the candidate molecule. A compound which promotes a decrease in the expression of the pathogenicity factor is considered useful in the invention; such a molecule may be used, for example, as a therapeutic to combat the pathogenicity of an infectious organism.

If desired, the effect of candidate compounds may, in the alternative, be measured at the level of polypeptide production using the same general approach and standard immunological techniques, such as Western blotting or immunoprecipitation with an antibody specific for a pathogenicity factor. For example, immunoassays may be

used to detect or monitor the expression of at least one of the polypeptides of the invention in a pathogenic organism. Polyclonal or monoclonal antibodies (produced as described above) which are capable of binding to such a polypeptide may be used in any standard immunoassay format (e.g., ELISA, Western blot, or RIA assay) to measure the level of the pathogenicity polypeptide. A compound which promotes a decrease in the expression of the pathogenicity polypeptide is considered particularly useful. Again, such a molecule may be used, for example, as a therapeutic to combat the pathogenicity of an infectious organism.

Alternatively, or in addition, candidate compounds may be screened for those which specifically bind to and inhibit a pathogenicity polypeptide of the invention. The efficacy of such a candidate compound is dependent upon its ability to interact with the pathogenicity polypeptide. Such an interaction can be readily assayed using any number of standard binding techniques and functional assays (e.g., those described in Ausubel et al., *supra*). For example, a candidate compound may be tested *in vitro* for interaction and binding with a polypeptide of the invention and its ability to modulate pathogenicity may be assayed by any standard assays (e.g., those described herein).

In one particular example, a candidate compound that binds to a pathogenicity polypeptide may be identified using a chromatography-based technique. For example, a recombinant polypeptide of the invention may be purified by standard techniques from cells engineered to express the polypeptide (e.g., those described above) and may be immobilized on a column. A solution of candidate compounds is then passed through the column, and a compound specific for the pathogenicity polypeptide is identified on the basis of its ability to bind to the pathogenicity polypeptide and be immobilized on the column. To isolate the compound, the column is washed to remove non-specifically bound molecules, and the compound of interest is then released from the column and collected. Compounds isolated by this method (or any other appropriate method) may, if desired, be further purified (e.g., by high performance liquid chromatography). In addition, these candidate compounds may be tested for their ability to render a pathogen less virulent (e.g., as described herein). Compounds isolated by this approach may also be used, for example, as therapeutics to treat or prevent the onset of a pathogenic

infection, disease, or both. Compounds which are identified as binding to pathogenicity polypeptides with an affinity constant less than or equal to 10 mM are considered particularly useful in the invention.

In yet another approach, candidate compounds are screened for the ability to inhibit the virulence of a *Pseudomonas* cell by monitoring the effect of the compound on the production of a phenazine (e.g., pyocyanin). According to one approach, candidate compounds are added at varying concentrations to a culture medium of pathogenic cells. Pyocyanin is then measured according to any standard method, for example, by monitoring its absorbance at 520 nm in acidic solution (Essar et al., *J. Bacteriol.* 172: 884, 1990). To maximize pyocyanin production in liquid culture for quantitation, cells may be cultured in a modified KA broth (King et al., *J. Lab. Clin. Med.* 44:301, 1954) by adding 100 μM $FeCl_3$. The level of pyocyanin production in the presence of the candidate compound is compared to the level measured in a control culture medium lacking the candidate molecule. A compound which promotes a decrease in the expression of a pyocyanin is considered useful in the invention; such a molecule may be used, for example, as a therapeutic to combat the pathogenicity of an infectious organism. Similar techniques may also be used to screen for other appropriate phenazines including, without limitation, pyorubin, aeruginosin A, myxin, and tubermycin A. Other phenazines are described in Turner and Messenger (*Advances In Microbial Physiology* 27:211-1275, 1986), Sorensen and Joseph (In: *Pseudomonas aeruginosa as an Opportunistic Pathogen*, Campa, M., ed., Plenum Press, N.Y., 1993), Ingram and Blackwood (*Advances in Applied Microbiology* 13: 267, 1970), and Gerber (In: CRC Handbook of Microbiology, Laskin, A.I., and Lechevalier, eds., 2nd edition, vol. 5, Chemical Rubber Co., Cleveland, Ohio, 1984, pp. 573-576).

Potential antagonists include organic molecules, peptides, peptide mimetics, polypeptides, and antibodies that bind to a nucleic acid sequence or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also include small molecules that bind to and occupy the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Other potential antagonists include antisense molecules.

Each of the DNA sequences provided herein may also be used in the discovery and development of antipathogenic compounds (e.g., antibiotics). The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide, or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for infection. In particular the molecules of the invention may be used: in the prevention of adhesion and colonization of bacteria to mammalian extracellular matrix proteins; to extracellular matrix proteins in wounds; to block mammalian cell invasion; or to block the normal progression of pathogenesis.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat a variety of bacterial infections.

Optionally, compounds identified in any of the above-described assays may be confirmed as useful in conferring protection against the development of a pathogenic infection in any standard animal model (e.g., the mouse-burn assay described herein) and, if successful, may be used as anti-pathogen therapeutics (e.g, antibiotics).

Test Compounds and Extracts

In general, compounds capable of reducing pathogenic virulence are identified from large libraries of both natural product or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous

methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their anti-pathogenic activity should be employed whenever possible.

When a crude extract is found to have an anti-pathogenic or anti-virulence activity, or a binding activity, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having anti-pathogenic activity. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for the treatment of pathogenicity are chemically modified according to methods known in the art.

Pharmaceutical Therapeutics and Plant Protectants

The invention provides a simple means for identifying compounds (including peptides, small molecule inhibitors, and mimetics) capable of inhibiting the pathogenicity or virulence of a pathogen. Accordingly, a chemical entity discovered to

have medicinal or agricultural value using the methods described herein are useful as either drugs, plant protectants, or as information for structural modification of existing anti-pathogenic compounds, e.g., by rational drug design. Such methods are useful for screening compounds having an effect on a variety of pathogens including, but not limited to, bacteria, viruses, fungi, annelids, nematodes, platyhelminthes, and protozoans. Examples of pathogenic bacteria include, without limitation, *Aerobacter*, *Aeromonas*, *Acinetobacter*, *Agrobacterium*, *Bacillus*, *Bacteroides*, *Bartonella*, *Bortella*, *Brucella*, *Calymmatobacterium*, *Campylobacter*, *Citrobacter*, *Clostridium*, *Cornyebacterium*, *Enterobacter*, *Escherichia*, *Francisella*, *Haemophilus*, *Hafnia*, *Helicobacter*, *Klebsiella*, *Legionella*, *Listeria*, *Morganella*, *Moraxella*, *Proteus*, *Providencia*, *Pseudomonas*, *Salmonella*, *Serratia*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Treponema*, *Xanthomonas*, *Vibrio*, and *Yersinia*.

For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Treatment may be accomplished directly, e.g., by treating the animal with antagonists which disrupt, suppress, attenuate, or neutralize the biological events associated with a pathogenicity polypeptide. Preferable routes of administration include, for example, subcutaneous, intravenous, interperitoneally, intramuscular, or intradermal injections which provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of an anti-pathogenic agent in a physiologically-acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's *Pharmaceutical Sciences* by E.W. Martin. The amount of the anti-pathogenic agent (e.g., an antibiotic) to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the type of disease and extensiveness of the disease. Generally, amounts will be in the range of those used for other agents used in the treatment of other microbial diseases, although in certain instances lower amounts will be needed because of the increased specificity of the compound. A compound is administered at a dosage that inhibits microbial proliferation. For example, for systemic administration a compound is

administered typically in the range of 0.1 ng - 10 g/kg body weight.

For agricultural uses, the compositions or agents identified using the methods disclosed herein may be used as chemicals applied as sprays or dusts on the foliage of plants. Typically, such agents are to be administered on the surface of the plant in advance of the pathogen in order to prevent infection. Seeds, bulbs, roots, tubers, and corms are also treated to prevent pathogenic attack after planting by controlling pathogens carried on them or existing in the soil at the planting site. Soil to be planted with vegetables, ornamentals, shrubs, or trees can also be treated with chemical fumigants for control of a variety of microbial pathogens. Treatment is preferably done several days or weeks before planting. The chemicals can be applied by either a mechanized route, e.g., a tractor or with hand applications. In addition, chemicals identified using the methods of the assay can be used as disinfectants.

Other Embodiments

In general, the invention includes any nucleic acid sequence which may be isolated as described herein or which is readily isolated by homology screening or PCR amplification using the nucleic acid sequences of the invention. Also included in the invention are polypeptides which are modified in ways which do not abolish their pathogenic activity (assayed, for example as described herein). Such changes may include certain mutations, deletions, insertions, or post-translational modifications, or may involve the inclusion of any of the polypeptides of the invention as one component of a larger fusion protein. Also, included in the invention are polypeptides that have lost their pathogenicity.

Thus, in other embodiments, the invention includes any protein which is substantially identical to a polypeptide of the invention. Such homologs include other substantially pure naturally-occurring polypeptides as well as allelic variants; natural mutants; induced mutants; proteins encoded by DNA that hybridizes to any one of the nucleic acid sequences of the invention under high stringency conditions or, less preferably, under low stringency conditions (e.g., washing at 2X SSC at 40°C with a probe length of at least 40 nucleotides); and proteins specifically bound by antisera of the

invention.

The invention further includes analogs of any naturally-occurring polypeptide of the invention. Analogs can differ from the naturally-occurring the polypeptide of the invention by amino acid sequence differences, by post-translational modifications, or by both. Analogs of the invention will generally exhibit at least 85%, more preferably 90%, and most preferably 95% or even 99% identity with all or part of a naturally-occurring amino acid sequence of the invention. The length of sequence comparison is at least 15 amino acid residues, preferably at least 25 amino acid residues, and more preferably more than 35 amino acid residues. Again, in an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence. Modifications include *in vivo* and *in vitro* chemical derivatization of polypeptides, e.g., acetylation, carboxylation, phosphorylation, or glycosylation; such modifications may occur during polypeptide synthesis or processing or following treatment with isolated modifying enzymes. Analogs can also differ from the naturally-occurring polypeptides of the invention by alterations in primary sequence. These include genetic variants, both natural and induced (for example, resulting from random mutagenesis by irradiation or exposure to ethanemethylsulfate or by site-specific mutagenesis as described in Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual* (2d ed.), CSH Press, 1989, or Ausubel et al., *supra*). Also included are cyclized peptides, molecules, and analogs which contain residues other than L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids.

In addition to full-length polypeptides, the invention also includes fragments of any one of the polypeptides of the invention. As used herein, the term "fragment," means at least 5, preferably at least 20 contiguous amino acids, preferably at least 30 contiguous amino acids, more preferably at least 50 contiguous amino acids, and most preferably at least 60 to 80 or more contiguous amino acids. Fragments of the invention can be generated by methods known to those skilled in the art or may result from normal protein processing (e.g., removal of amino acids from the nascent polypeptide that are not required for biological activity or removal of amino acids by alternative mRNA splicing

or alternative protein processing events).

Furthermore, the invention includes nucleotide sequences that facilitate specific detection of any of the nucleic acid sequences of the invention. Thus, for example, nucleic acid sequences described herein or fragments thereof may be used as probes to hybridize to nucleotide sequences by standard hybridization techniques under conventional conditions. Sequences that hybridize to a nucleic acid sequence coding sequence or its complement are considered useful in the invention. Sequences that hybridize to a coding sequence of a nucleic acid sequence of the invention or its complement and that encode a polypeptide of the invention are also considered useful in the invention. As used herein, the term "fragment," as applied to nucleic acid sequences, means at least 5 contiguous nucleotides, preferably at least 10 contiguous nucleotides, more preferably at least 20 to 30 contiguous nucleotides, and most preferably at least 40 to 80 or more contiguous nucleotides. Fragments of nucleic acid sequences can be generated by methods known to those skilled in the art.

The invention further provides a method for inducing an immunological response in an individual, particularly a human, which includes inoculating the individual with, for example, any of the polypeptides (or a fragment or analog thereof or fusion protein) of the invention to produce an antibody and/or a T cell immune response to protect the individual from infection, especially bacterial infection (e.g., a *Pseudomonas aeruginosa* infection). The invention further includes a method of inducing an immunological response in an individual which includes delivering to the individual a nucleic acid vector to direct the expression of a polypeptide described herein (or a fragment or fusion thereof) in order to induce an immunological response.

The invention also includes vaccine compositions including the polypeptides or nucleic acid sequences of the invention. For example, the polypeptides of the invention may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example, by blocking the production of phenazines. The invention therefore includes a vaccine formulation which includes an immunogenic recombinant polypeptide of the invention together with a suitable carrier.

The invention further provides compositions (e.g., nucleotide sequence probes), polypeptides, antibodies, and methods for the diagnosis of a pathogenic condition.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.

Other embodiments are within the scope of the claims.

What is claimed is: